A Neutrosophic Framework for Assessing Ethical Considerations in Autonomous Systems

ANTONIOS PARASKEVAS
Department of Applied Informatics,
University of Macedonia,
156 Egnatias Str., Thessaloniki, 54636,
GREECE

Abstract: - As autonomous systems become an integral part of everyday life, their decision-making processes bring forth significant ethical challenges. To fully harness the potential of these systems, their design and development must adhere to core ethical principles and societal values. This calls for the adoption of innovative methodologies and approaches that embed ethical considerations into the creation of autonomous intelligent systems. Building on prior work evaluating algorithmic ethics through Neutrosophic Logic, this study introduces the Neutrosophic Ethical Integrity Score (NEIS). The NEIS accounts for the complexities and uncertainties inherent in ethical decision-making by integrating three components: Truth (T), representing the ethical validity of actions; Falsehood (F), quantifying potential negative outcomes; and Indeterminacy (U), addressing uncertainties in ethical judgments. Through a case study of autonomous vehicles, this framework demonstrates its capability to assess and address ethical dilemmas. The findings underline its potential to identify and mitigate ethical risks, offering a pathway to the responsible advancement and deployment of autonomous technology.

Key-Words: - autonomous systems, ethical decision-making, neutrosophic logic, neutrosophic ethical integrity score, responsible AI, uncertainty in ethics.

Received: May 21, 2024. Revised: December 15, 2024. Accepted: January 17, 2025. Published: April 15, 2025.

1 Introduction

The incorporation of artificial intelligence (AI) into autonomous systems is one of the most significant technical developments of the twenty-first century, [1]. Autonomous systems, such as self-driving cars, unmanned aerial vehicles (UAVs), and industrial robots, rely largely on AI to accomplish tasks that need little to no human participation. Although they provide undeniable benefits, they are accompanied by concerns. It is critical that such machines act as a positive element of our society, which implies that they must behave ethically.

Algorithms play an important role in all computing systems, especially autonomous ones. In many respects, algorithms—whether those implemented in the autonomous system itself or those utilized for its learning and training—constitute the "mind" of the autonomous system. Algorithms have the potential to exacerbate prejudice, reduce transparency, and undermine accountability. Many variables contribute to the difficulty of determining an algorithm's potential and actual ethical effect. Identifying the influence of human subjectivity in algorithm design and configuration frequently necessitates an investigation of long-term, multi-user development processes. In recent years, scholars have sought to identify and categorize the ethical issues that algorithms raise, as well as the remedies presented in the relevant literature. The conceptual map developed by [2] remains a useful foundation for analyzing the current discussion on the ethics of algorithms.

Driven by the observation of inevitable uncertain knowledge when algorithms draw conclusions from the data they process using inferential statistics and/or machine learning techniques, scholars in [3] re-mapped the debate introducing the concept of indeterminacy into all aspects of the ethical map. In this context, they introduced a new neutrosophic methodological framework for evaluating ethical correctness (truth), ethical breaches (falsehood), and uncertainty (indeterminacy) in algorithms, offering a more comprehensive approach to ethical assessment. Neutrosophic logic has made its way into various applications, such as evaluation and decision-making processes, due to their beneficial approach used when one is unsure about the accuracy of the parameters / qualitative grades assigned to the members of the universal set, [4], [5].

While in [3] scholars introduced neutrosophic logic as a method for addressing ethical uncertainty in algorithms, this paper considerably enhances the state of the art by suggesting the Neutrosophic Ethical Integrity Score (NEIS) framework. The NEIS framework offers a mathematical approach to systematically evaluate the truth, falsity, and indeterminacy components of ethical decision-making in autonomous systems.

Assessing the ethical implications of autonomous systems is a challenging task due to the inherent complexities and uncertainties in their operations. Traditional frameworks frequently struggle to account for the diverse and multifaceted nature of the ethical dilemmas these technologies encounter, particularly in situations involving conflicting values or ambiguous circumstances, [6], [7], [8]. This highlights the need for novel approaches capable of navigating the intricate ethical issues that arise in autonomous decision-making.

In response to the aforementioned, the current study suggests a Neutrosophic framework for evaluating ethical issues in autonomous systems. This neutrosophic mathematical framework, first discussed in [3], was utilized as a tool in order to quantify algorithmic ethics in the area of healthcare. By using neutrosophic logic theory, this model attempts to measure ethical integrity in different features of ethical concerns that are encountered in autonomous systems and specifically in autonomous driving. It should be mentioned that the novelty of the current study lies in utilizing the NEIS framework to analyze real-world datasets from a unique ethical standpoint. Unlike previous applications of such datasets, which were solely concerned with technical performance indicators like detection accuracy or collision rates, this study incorporates the results into a comprehensive ethical evaluation framework. Furthermore, our literature review has revealed that there are very few related studies in this particular scientific area. This highlights the need for additional research, given the importance of this topic in many facets of our society.

The rest of the paper is organized as follows: Section 2 summarizes the related literature on ethical issues in autonomous systems and discusses the research gap identified regarding the lack of related studies in integrating the inherent uncertainty and indeterminate nature of ethical decision-making in such systems. In Section 3 we explain the Neutrosophic framework, including its fundamental definitions and equations. The next section demonstrates the applicability of the suggested method to real-world case studies analyzing ethical concerns encountered by autonomous vehicles. The article closes with the final conclusions and some hints for further research, contained in its last section.

2 Literature Review

Autonomous systems refer to technologies capable of performing tasks and making decisions without direct human intervention. These systems leverage advanced algorithms, artificial intelligence (AI), machine learning, and sensor technologies to analyze data, interpret environments, and execute actions in real time. Examples include self-driving vehicles [9], unmanned aerial vehicles (drones) [10], and robotic process automation in manufacturing and service industries [11], [12]. The deployment of autonomous systems entails important ethical concerns, which must be carefully studied in order to ensure responsible development and incorporation into society. The implications include accountability [13], [14], [15], transparency [16], [17], [18], prejudice [19], [20], [21], safety [22], [23], [24], privacy [25], [26], [27], [28], and the alignment of these systems with human values.

In the human and machine contexts, [29] define accountability as the ability to determine whether a system's decision was made in accordance with procedural and substantive standards, as well as to hold someone accountable when the standards are not met. Accountability becomes a difficult issue in autonomous driving, mostly because of the different activities involved (e.g., perception, planning, controls, system management, among others) that require inputs from multiple stakeholders; this can result in responsibility gaps. As autonomous vehicles rely more and more on deep neural networks processing visual streams [30], it is of critical importance to study the explainability of driving models from a computer vision perspective. The term explainability often co-occurs with the concept of interpretability, [31]. In particular, interpretability encompasses two key concepts: model transparency and post-hoc interpretability. Increasing model transparency equates to a better understanding of how the model works. For example, [32] describe that a decision model is transparent if its decision-making process can be immediately comprehended without any further information. If an external tool or model is utilized to explain the decision-making process, the offered explanation is not transparent. For example, in autonomous driving, transparency is critical because decisions such as braking or swerving must be immediately comprehensible to both passengers and, if necessary, external evaluators. One of the primary ethical challenges associated with autonomous systems is the potential for bias in AI algorithms. Biased algorithms in crucial applications, like as autonomous cars, might lead to unfair or dangerous outcomes, disproportionately hurting specific populations. To ensure fairness, AI systems must be rigorously tested, validated, and monitored on an ongoing basis to discover and minimize prejudice, [33]. Safety is a crucial aspect of autonomous driving, as it assists drivers and reduces the likelihood of accidents. This criterion is primarily met by seven important tasks: road detection [34], lane detection [35], vehicle detection [36], pedestrian detection [37], drowsiness detection [38], collision avoidance [39], and traffic sign detection [40]. According to the report about autonomous driving technology from the National Science & Technology Council (NSTC) and the United States Department of Transportation (USDOT) [41], the issue of privacy in autonomous vehicles is related to capturing a massive amount of sensor data from the environment. For example, the pedestrian's face and the license plate captured by the vehicle's camera should be masked as soon as possible, [42]. Furthermore, who owns the driving data is also an important issue, which requires the system's support for data access, storage, and communication, [43]. Lastly, the alignment problem, ensuring AI systems align with human values and intentions, is considered a critical scientific challenge. This alignment involves embedding principles like fairness, inclusivity, and respect for autonomy into algorithmic design and operation. The 'Moral Machine' project gathered feedback from millions of individuals on the moral trade-offs faced by autonomous vehicles, providing valuable data for bottom-up initiatives, [44]. Despite its large size, the study's findings were inconclusive. It indicated a collection of noisy preferences in this area, some clear inclinations (for example, valuing more lives over fewer lives), and some ethical variance between cultures, including a proclivity to assign more ethical weight to the lives of higher-status persons in impoverished countries. The study concerns the consistency of moral ideas and perspectives across communities, as well as the efficacy of an empirical approach to value selection when some opinions, such as those related to social position, may not be representative.

Despite the proliferation of studies addressing ethical implications in autonomous systems, current frameworks fail to adequately incorporate the inherent ambiguity and ambiguous nature of ethical decision-making in such systems. This article aims to bridge this gap by developing a neutrosophic mathematical framework for evaluating ethical integrity in autonomous systems. This methodology offers a nuanced approach, quantitatively evaluating the truth, falsity, and uncertainty elements of ethical issues, providing a fresh perspective in ethical AI research.

3 Mathematical Framework

In this section, we will briefly present the mathematical background discussed in [3].

A logic in which each proposition is estimated to have the percentage of truth in a subset T, the percentage of indeterminacy in a subset I, and the percentage of falsity in a subset F, where T, I, F are defined above, is called Neutrosophic Logic, [45].

In the following, T, I, and F, called neutrosophic components, will represent the truth value, indeterminacy value, and falsehood value respectively referring to neutrosophy, neutrosophic logic, and neutrosophic set.

In this framework, a formula φ is characterized by a triplet of truth values, called the neutrosophic value defined as [46]:

$$\mathrm{NL}(\varphi) = (\mathrm{T}(\varphi), \mathrm{I}(\varphi), \mathrm{F}(\varphi)) \text{ where } (\mathrm{T}(\varphi), \mathrm{I}(\varphi), \mathrm{F}(\varphi)) \subset \|{-0}, 1+\|^3 \quad (1)$$

For each algorithmic ethical criterion C_i $(i = 1,2 \ldots, n)$, we can represent its performance by utilizing the concept of neutrosophic set in a similar way as given in (1):

$$C_i = \{(x, T_i(x), I_i(x), F_i(x)) | x \in U\} \quad (2)$$

where U : the universe of discourse (e.g. algorithm outputs)

Definition 1 [3]. The Neutrosophic Ethical Integrity Score (NEIS) for a given ethical criterion $\boldsymbol{i}$ is defined as:
$NEIS_i : \mathbb{R}^3 \rightarrow \mathbb{R}$ where $\mathbb{R}^3$ represents the three-dimensional space of the components of ethical evaluation, specifically truth, falsehood, and indeterminacy.

The $\boldsymbol{NEIS_i}$ function is given as follows:

$$NEIS_i = T_i - F_i + I_i \quad (3)$$

Definition 2 [3]. Let n be the number of criteria for assessing algorithmic ethics. The Overall NEIS can be defined as:

Overall NEIS (OvNEIS) = $\sum_{i=1}^{n} w_i * NEIS_i$ (4)

where w_i is the weight assigned to algorithmic ethical criterion i such as $\sum_{i=1}^{n} w_i = 1$.

4 Results

In this section, we will study and apply the proposed framework to a real-world case study involving autonomous driving. The scenario we are examining has been well-documented in numerous literature sources [13], [47], [48], [49], [50], [51], but from a different perspective. Each study has presented frameworks or findings relevant to the ethical decision-making process for autonomous vehicles (AVs) in scenarios similar to the one we are studying. However, in our framework, we will quantify the importance of ethical criteria. Based on the results obtained, the autonomous system (AV in our case) will make the most ethical decision. This introduces a structured, quantitative approach to handling ethical indeterminacy, which is largely absent in existing AV ethics frameworks.

4.1 Scenario

An autonomous vehicle (AV) navigates a crowded metropolitan environment with frequent pedestrian crossings, bikers, and changing road conditions. The AV is intended to prioritize safety and make ethical judgments in real-time, utilizing data from sensors and algorithms that identify and respond to things in its path. The AV is presented with two options:

1. Sudden Stop: A pedestrian unexpectedly enters the road. The AV can perform an emergency stop to avoid the pedestrian, but this sudden move may create discomfort for the passengers and perhaps result in a rear-end accident with the car behind.
2. Controlled Evasion: The AV can execute a controlled maneuver to avoid the pedestrian without arriving at a complete stop, although this may put it closer to a bike in the next lane.

4.2 Data

The data used in this study includes a combination of real-world sensor logs, simulation data, and public datasets relevant to autonomous vehicle (AV) ethical performance. More specifically, The KITTI Vision Benchmark Suite [52] and the Waymo Open Dataset [53] were used to analyze pedestrian identification accuracy and collision rates. Demographic bias in detection accuracy is evaluated using methodologies inspired by datasets such as COMPAS, [28]. Transparency and interpretability scores are derived from simulation environments like the CARLA Autonomous Driving Simulator, [54]. Finally, privacy-related metrics are informed by research on privacy-preserving machine learning, [55]. Also, Open Mobility Data from the U.S. Department of Transportation (DOT) provides datasets related to AV safety, crash rates, and privacy regulations in real-world deployment contexts.

4.3 Application

In Table 1, the data used and their relevance to each component of the NEIS framework are depicted:

Table 1. Datasets used in the scenario

Metric	**Dataset**	**Description**	**Purpose of NEIS framework**
Pedestrian Detection Accuracy	Waymo Open Dataset, KITTI Vision Benchmark Suite	High-resolution sensor data with labelled objects, including pedestrians and cyclists.	Assesses **Truth (T)** by evaluating AV's accuracy in detecting pedestrians.
Collision Rate	Waymo Open Dataset	Incident records and collision statistics in various driving environments.	Contributes to **Falsehood (F)** by quantifying AV's safety performance and frequency of collisions.
Demographic Bias	COMPAS	Data and methodology used to analyse demographic disparities in AV detection rates and ethical decisions.	Evaluate bias **Mitigation** within **Truth (T)** and **Falsehood (F)** by measuring fairness and inclusivity.
Transparency	CARLA Autonomous Driving Simulator	Simulation data enables transparency assessments and interpretability metrics for AV decisions.	Quantifies **Transparency** in **Truth (T)** by measuring the AV's decision-making clarity.
Privacy Anonymization	U.S. DOT Mobility Data	Datasets and methods for anonymizing sensitive AV data (e.g., video, location)	Addresses **Indeterminacy (U)** by ensuring privacy and ethical data.

Next, we will describe how we obtain, for example, the Truth (T) value for pedestrian detection accuracy from dataset Waymo Open Dataset and KITTI Vision Benchmark Suite.

We count:
True Positives (TP): Instances where pedestrians are correctly identified.
True Negatives (TN): Instances where non-pedestrians are correctly identified.
False Positives (FP): Instances where non-pedestrians are incorrectly identified as pedestrians.
False Negatives (FN): Instances where pedestrians are incorrectly identified as non-pedestrian

Then accuracy can be calculated using the formula:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

If we are focusing only on the accuracy of True positives:

$$\text{PedDetAccuracy} = \frac{TP}{TP+TN} \tag{6}$$

From the dataset, we obtain:
True Positives (TP) = 900
True Negatives (TN) = 800
False Positives (FP) = 50
False Negatives (FN) = 150
Based on the above data and from Eq. (6) we have:
PedDetAccuracy = 0.857.
This accuracy value is used as the truth value (T) of pedestrian detection.
Next, we define:

$$\text{CollisionRate} = \frac{Collisions}{Total\ Encounters} \tag{7}$$

From dataset, we have Total encounters = 1.000 and Collisions = 20.

So, from Eq. (7), we obtain CollisionRate =0.02.

The above value represents the Falsehood (F) component for collision rate.

By following the same logic with other metrics as well we obtain Table 2.

Table 2. Results of metrics

Metric	T	F	I	Weight
Pedestrian Detection	0.857	0	0	0.25
Collision Rate	0	0.02	0.05	0.20
Demographic Bias	0.75	0.25	0.05	0.20
Transparency	0.70	0	0	0.15
Privacy Anonymization	0	0	0.90	0.20

Remark 1: The weights assigned to each measure in the NEIS computation indicate their respective relevance in the ethical assessment of autonomous systems. These weights were calculated using a mix of literature analysis, domain expertise, and practical factors in autonomous vehicle operations. For example, pedestrian identification accuracy was given greater weight (0.25) since it has a direct influence on safety and ethical decision-making, which is a top priority in autonomous systems. Similarly, privacy anonymization, which deals with sensitive data management, received a significant weight (0.20) to reflect rising public concerns about data privacy and ethical data usage.

Remark 2: Privacy anonymization is linked with indeterminacy component (I) because Privacy anonymization is linked to the Indeterminacy (U) component because it expresses ethical concerns about how private data will be managed, secured, and possibly reused.

Collision rate contributes to Indeterminacy (I) except Falsehood (F) as well, due to uncertainty in high-risk or unpredictable environments.

Uncertainty in fairness across different demographic groups contributes to Indeterminacy (I) in metric Demographic Bias.

The weights assigned to each metric in the NEIS calculation reflect the importance or priority given to each ethical criterion in the AV's decision-making process.

Now we can calculate the Overall Neutrosophic Ethical Integrity Score (OvNEIS) by Eq. (4) and data given in Table 2.

OvNEIS = (0.4693, 0.18, 0.054)

Given the above value, we can indicate that we are encountered with an ethically trustworthy AV, with small risks of damage and a few areas that need further attention to reduce ambiguity in complicated ethical scenarios.

5 Conclusion

The ethical implications of autonomous vehicles (AVs) are critical because they address essential concerns such as safety, responsibility, and moral decision-making. AVs function without direct human direction, posing complicated considerations concerning accountability in accidents and the ethical frameworks that guide their key judgments, [56], [57]. Traditional frameworks may not sufficiently address the multidimensional nature of ethical challenges posed by new technologies, particularly in scenarios with opposing values and unclear situations.

In this study, we employed a neutrosophic metric, namely the Overall Neutrosophic Ethical Integrity Score, for evaluating ethical integrity in AVs. This is achieved by evaluating the truth, falsity, and uncertainty elements of ethical issues, thus suggesting the application of a quantitative mathematical framework that has the capacity to capture the complex and nuanced nature of ethical dilemmas faced by AVs. Incorporating uncertainty into the ethical evaluation of autonomous systems is a crucial step toward addressing the real-world challenges these technologies present. In this way, the suggested model can also aid in developing guidelines for ethical programming and policy-making, ultimately promoting safer and more socially responsible AVs.

Finally, addressing environmental ethics will guarantee that AV technology is consistent with sustainability objectives, increasing social trust and adoption. For non-technical audiences, the findings demonstrate how developing technologies may be driven by ethical standards to assure safety, justice, and transparency in everyday applications such as autonomous driving. By connecting technological measures with moral concerns, this research highlights technology's ability to positively improve society while adhering to common ethical norms.

The proposed Neutrosophic Ethical Integrity Score (NEIS) framework opens up several avenues for future research. In this perspective, one could examine its applicability to other fields, such as AI-driven diagnostic tools, or financial systems. Furthermore, integrating the NEIS framework with emerging AI technologies such as large language models might assure ethical adherence in systems capable of making complicated decisions. Real-time adaptations of the NEIS framework to dynamic operational contexts remain an unresolved problem, particularly in high-stake environments. Lastly, our proposed model could be utilized for the enhancement of the NEIS framework to encompass more varied ethical factors, such as including environmental effects and emotional intelligence in decision-making.

In summary, the NEIS framework provides a flexible and effective method for addressing the ethical dilemmas faced by autonomous systems. By accounting for the inherent uncertainty and ethical risks involved in decision-making, this framework ensures that these technologies evolve in a manner that aligns with fundamental ethical principles.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work the author used "EditMyEnglish" in order to to improve the clarity and grammar of the manuscript's English language. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References:

[1] Nguyen Q., Beeram, D., Li, Y., Brown, S.J., and Yuchen, N., *Expert matching through workload intelligence,* ed: Google Patents, 2022.

[2] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., and Floridi, L, The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2016, 2053951716679679. https://doi.org/10.1177/2053951716679679.

[3] Paraskevas, A., and Smarandache, F., Introduction to algorithmic ethics under a neutrosophic framework: Re-mapping the debate. *Neutrosophic Sets and Systems*, *75*, 2024, 302-308.

[4] Voskoglou, M. G. (2023). Neutrosophic Assessment and Decision Making. *Equations*, *3*, 68-72.

[5] Voskoglou, M. G. (2023). Decision Making in Neutrosophic Environment. *Proof*, *3(1)*, 1-7. DOI: 10.37394/232020.2023.3.1.

[6] Veale, M., Van Kleek, M., & Binns, R., Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In *Proceedings of the 2018 chi conference on human factors in computing systems*, 1-14. https://doi.org/10.1145/3173574.3174014.

[7] Hajian S. and Domingo-Ferrer, J., Direct and indirect discrimination prevention methods. *In Discrimination and privacy in the information society,* Bart Custers, Toon Calders, Bart Schermer, and Tal Zarsky (Eds.). Springer, Berlin, Heidelberg, 2012, pp.241–254. https://doi.org/10.1007/978-3-642-30487-3_13.

[8] Pedreshi, D., Ruggieri, S., and Turini, F., Discrimination-aware Data Mining. In Proceedings of the *14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, Las Vegas, USA, 560-568. https://doi.org/10.1145/1401890.1401959.

[9] El-Sebah, M. I. A., Syam, F. A., Sweelem, E. A., & El-Sotouhy, M. M. (2023). A Proposed

Controller for an Autonomous Vehicles Embedded System. *WSEAS Transactions on Circuits and Systems*, vol.22, pp.1-9. https://doi.org/10.37394/23201.2023.22.1.

[10] Yauri, R., Fernandez, S., & Aquino, A. (2024). Control of Autonomous Aerial Vehicles to Transport a Medical Supplies. *WSEAS Transactions on Systems*, vol.23, pp.73-81. https://doi.org/10.37394/23202.2024.23.8.

[11] Vaisi, B. A. H. A. R. E. H., Farughi, H., & Raissi, S. (2020). Multiobjective optimal model for task scheduling and allocation in a two machines robotic cell considering breakdowns. *WSEAS Transactions on Information Science and Applications*, vol.17, pp.1-8. https://doi.org/10.37394/23209.2020.17.a025103-920.

[12] Ramirez, J. A. R., Franco, E. M., & Varela, D. T. (2017). Human-machine interfaces using fuzzification of facial movements to remotely control robots by means of XMPP internet protocol. *WSEAS Transactions on Communications*, vol.16, pp.94-106. https://doi.org/10.1051/matecconf/201712504020.

[13] Gogoll, J., and Müller, J. F., Autonomous cars: in favor of a mandatory ethics setting. *Science and Engineering Ethics*, vol.23, 2017, pp.681-700. https://doi.org/10.1007/s11948-016-9806-x.

[14] Verdiesen, I., Santoni de Sio, F., and Dignum, V., Accountability and control over autonomous weapon systems: a framework for comprehensive human oversight. *Minds and Machines*, 31(1), 2021, pp.137-163. https://doi.org/10.1007/s11023-020-09532-9.

[15] Omeiza, D., Web, H., Jirotka, M., and Kunze, L., Towards accountability: Providing intelligible explanations in autonomous driving. In *2021 IEEE Intelligent Vehicles Symposium (IV)*, pp. 231-237. IEEE. https://doi.org/10.1109/iv48863.2021.9575917.

[16] Miller, T., Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, vol.267, 2019, pp.1-38. https://doi.org/10.1016/j.artint.2018.07.007.

[17] Bhaskara, A., Skinner, M., and Loft, S., Agent transparency: A review of current theory and evidence. *IEEE Transactions on Human-Machine Systems*, vol.50 (3), 2020, pp.215-224. https://doi.org/10.1109/thms.2020.2965529.

[18] Andrada, G., Clowes, R. W., and Smart, P. R., Varieties of transparency: Exploring agency within AI systems. *AI & society*, vol.38 (4), 2023, pp.1321-1331. https://doi.org/10.1007/s00146-021-01326-6.

[19] Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S., Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, vol. 366(6464), 2019, pp.447-453. https://doi.org/10.1126/science.aax2342.

[20] Chen, Z., Collaboration among recruiters and artificial intelligence: removing human prejudices in employment. *Cognition, Technology & Work*, vol.25 (1), 2023, pp.135-149. https://doi.org/10.1007/s10111-022-00716-0.

[21] Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., and Lowry, P. B., Collaborating with technology-based autonomous agents: Issues and research opportunities. *Internet Research*, vol.30 (1), 2020, 1-18. https://doi.org/10.1108/intr-12-2019-0503.

[22] Alves, E. E., Bhatt, D., Hall, B., Driscoll, K., Murugesan, A., and Rushby, J., *Considerations in assuring safety of increasingly autonomous systems* (No. NASA/CR-2018-220080), 2018.

[23] Burton, S., Habli, I., Lawton, T., McDermid, J., Morgan, P., and Porter, Z., Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective. *Artificial Intelligence*, vol.279, 2020, 103201. https://doi.org/10.1016/j.artint.2019.103201.

[24] Ivanov, R., Carpenter, T. J., Weimer, J., Alur, R., Pappas, G. J., and Lee, I., Verifying the safety of autonomous systems with neural network controllers. *ACM Transactions on Embedded Computing Systems (TECS)*, vol.20 (1), 2020, 1-26. https://doi.org/10.1145/3419742.

[25] Lera, F. J. R., Llamas, C. F., Guerrero, Á. M., & Olivera, V. M., Cybersecurity of robotics and autonomous systems: Privacy and safety. *Robotics-legal, ethical and socioeconomic impacts*, 2017. https://doi.org/10.5772/intechopen.69796.

[26] Such, J. M., Privacy and autonomous systems. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial*

Intelligence, IJCAI-2017, pp.4761-4767. https://doi.org/10.24963/ijcai.2017/663.

[27] Karlin, J., Forrest, S., and Rexford, J., Autonomous security for autonomous systems. *Computer Networks*, vol.52 (15), 2008, pp.2908-2923. https://doi.org/10.1016/j.comnet.2008.06.012.

[28] Li, Y., Tao, X., Zhang, X., Liu, J., & Xu, J., Privacy-preserved federated learning for autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, vol.23 (7), 2021, 8423-8434. https://doi.org/10.1109/tits.2021.3081560.

[29] Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Schieber, S., Waldo, J., Weinberger, D., and Wood, A., Accountability of AI under the law: The role of explanation, arXiv preprint arXiv:1711.01134, 2017. https://doi.org/10.2139/ssrn.3064761.

[30] Janai, J., Güney, F., Behl, A., and Geiger, A. Computer vision for autonomous vehicles: Problems, datasets and state of the art. *Foundations and Trends® in Computer Graphics and Vision*, vol.12 (1–3), 2020, 1-308. https://doi.org/10.1561/0600000079.

[31] Zablocki, É., Ben-Younes, H., Pérez, P., and Cord, M., Explainability of deep vision-based autonomous driving systems: Review and challenges. *International Journal of Computer Vision*, vol.130 (10), 2022, 2425-2452. https://doi.org/10.1007/s11263-022-01657-x.

[32] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D., A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, vol.51 (5), 2018, pp.1-42. https://doi.org/10.1145/3236009.

[33] Kondam, A., and Yella, A., Artificial Intelligence and the Future of Autonomous Systems. *Innovative Computer Sciences Journal*, vol.9(1), 2023.

[34] Harish, H., and Murthy, A. S., Identification of lane lines using advanced machine learning. In *2022 8th international conference on advanced computing and communication systems (ICACCS)*, IEEE, Vol. 1, pp.1094-1099. https://doi.org/10.1109/icaccs54159.2022.9785221.

[35] Du, X., and Tan, K. K., Comprehensive and practical vision system for self-driving vehicle lane-level localization. *IEEE transactions on image processing*, vol.25 (5), 2016, pp.2075-2088. https://doi.org/10.1109/tip.2016.2539683.

[36] Ashouri, M.R, Nahvi, A. and Azadi, S. Time delay analysis of vehicle handling variables for near-crash detection of drowsy driving using a bus driving simulator," *in Proc. 6th RSI Int. Conf. Robot. Mechatronics (IcRoM)*, Oct. 2018, pp.243–249. https://doi.org/10.1109/icrom.2018.8657519.

[37] Ravula, A. K., Assisting visually impaired people by computer vision-a smart eye. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol.12 (11), 2021, pp.2118-2124.

[38] Picot, A., Charbonnier, S., and Caplier, A., On-line detection of drowsiness using brain and visual information. *IEEE Transactions on systems, man, and cybernetics-part A: systems and humans*, vol.42 (3), 2011, pp.764-775. https://doi.org/10.1109/tsmca.2011.2164242.

[39] Deng, R., Di, B., and Song, L., Cooperative collision avoidance for overtaking maneuvers in cellular V2X-based autonomous driving. *IEEE Transactions on Vehicular Technology*, vol.68 (5), 2019, pp.4434-4446. https://doi.org/10.1109/tvt.2019.2906509.

[40] Lee, H. S., and Kim, K., Simultaneous traffic sign detection and boundary estimation using convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, vol.19 (5), 2018, pp.1652-1663. https://doi.org/10.1109/tits.2018.2801560.

[41] U.S. Department of Transportation. (2020). *Ensuring American leadership in automated vehicle technologies: Automated vehicles 4.0,* [Online]. https://www.transportation.gov/av/4 (Accessed Date: October 26, 2024).

[42] Taghavi, S., and Shi, W. (2020, November). EdgeMask: An edge-based privacy preserving service for video data sharing. In *2020 IEEE/ACM Symposium on Edge Computing (SEC),* IEEE, pp.382-387. https://doi.org/10.1109/sec50012.2020.00056.

[43] Ren, K., Wang, Q., Wang, C., Qin, Z., and Lin, X., The security of autonomous driving: Threats, defences, and future directions. *Proceedings of the IEEE*, vol.108 (2), 2019, pp.357-372. https://doi.org/10.1109/jproc.2019.2948775.

[44] Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A.and Rahwan, I., The moral machine experiment. *Nature*, vol.563 (7729), 2018, pp.59-64. https://doi.org/10.1038/s41586-018-0637-6.

[45] Smarandache, F., Neutrosophic logic-a generalization of the intuitionistic fuzzy logic. *Multispace & multistructure. Neutrosophic transdisciplinarity (100 collected papers of science)*, vol.4, 2010, 396. https://doi.org/10.2139/ssrn.2721587.

[46] Smarandache, F., *A unifying field in Logics: Neutrosophic Logic*. In *Philosophy*, pp.1-141. American Research Press, 1999.

[47] Bonnefon, J. F., Shariff, A., and Rahwan, I., The social dilemma of autonomous vehicles. *Science*, vol.352 (6293), 2016, pp.1573-1576. https://doi.org/10.1126/science.aaf2654.

[48] Papadimitriou, E., Farah, H., van de Kaa, G., De Sio, F. S., Hagenzieker, M., and van Gelder, P.,Towards common ethical and safe 'behaviour'standards for automated vehicles. *Accident Analysis & Prevention*, vol.174, 2022, 106724. https://doi.org/10.1016/j.aap.2022.106724.

[49] Lim, H. S. M., and Taeihagh, A., Algorithmic decision-making in AVs: Understanding ethical and technical concerns for smart cities. *Sustainability*, vol.11 (20), 2019, 5791. https://doi.org/10.3390/su11205791.

[50] Martinho, A., Herber, N., Kroesen, M., and Chorus, C., Ethical issues in focus by the autonomous vehicles industry. *Transport reviews*, vol.41 (5), 2021, pp.556-577. https://doi.org/10.1080/01441647.2020.1862355.

[51] Fossa, F., *Ethics of driving automation: Artificial agency and human values* (Vol. 65). Springer Nature, 2023.

[52] Geiger, A., Lenz, P., Stiller, C., and Urtasun, R., Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, vol.32 (11), 2013, pp.1231-1237. https://doi.org/10.1177/0278364913491297.

[53] Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., et al.: Scalability in perception for autonomous driving: Waymo Open Dataset. *In: CVPR (2020)*. https://doi.org/10.1109/cvpr42600.2020.00252.

[54] Osiński, B., Miłoś, P., Jakubowski, A., Zięcina, P.,Martyniak, M., Galias, C., and Michalewski, H., CARLA Real Traffic Scenarios--novel training ground and benchmark for autonomous driving. *arXiv preprint arXiv:2012.*, 2020, 11329.

[55] Coeckelbergh, M., Responsibility and the moral phenomenology of using self-driving cars. *Applied Artificial Intelligence*, vol.30 (8), 2016, 748-757. https://doi.org/10.1080/08839514.2016.1229759.

[56] Arfini, S., Spinelli, D., and Chiffi, D., Ethics of self-driving cars: A naturalistic approach. *Minds and Machines*, vol.32 (4), 2022, pp.717-734. https://doi.org/10.1007/s11023-022-09587-5.

[57] Nyholm, S., The ethics of crashes with self-driving cars: A roadmap, I. *Philosophy Compass*, vol.13 (7), 2018. https://doi.org/10.1111/phc3.12507.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The author contributed to all stages of the research from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The author has no conflicts of interest to declare.