AFTEA Framework for Supporting Dynamic Autonomous Driving Situation

Subi Kim , Jieun Kang , Yongik Yoon *

Posted Date: 2 August 2024

doi: 10.20944/preprints202408.0126.v1

Keywords: Contextual Understanding in Dynamic Environments; Situation Awareness; Trustworthy and Sustainable AI; AFTEA

 doi:10.20944/preprints202408.0126.v1

Article

AFTEA Framework for Supporting Dynamic Autonomous Driving Situation

Subi Kim [1,†], Jieun Kang [1,†] and Yongik Yoon [2,*]

[1] Department of IT Engineering, Sookmyung Women's University, 100, Cheongpa-ro 47-gil, Yongsan-gu, Seoul 04310, South Korea; sbkimdbr@sookmyung.ac.kr (S.K.); kje9209@sookmyung.ac.kr (J.K.)

[2] Artificial Intelligence Engineering, Sookmyung Women's University, 100, Cheongpa-ro 47-gil, Yongsan-gu, Seoul 04310, South Korea

* Correspondence: yiyoon@sookmyung.ac.kr

† These authors contributed equally to this work.

Abstract: The accelerated development of AI technology has brought about revolutionary changes in various fields of society. Recently, it has been emphasized that Fairness, Accountability, Transparency, Ethics (Explainability) (FATE) should be considered to support the reliability and validity of AI-based decision-making. However, in the case of autonomous driving technology, which is directly related to human life and requires real-time adaptation and response to various changes and risks in the real world, environmental adaptability must be considered in a more comprehensive and converged manner. In order to derive definitive evidence for each object in a convergent autonomous driving environment, it is necessary to transparently collect and provide various types of road environment information for driving objects and driving assistance, and to construct driving technology that is adaptable to various situations by considering all uncertainties in the real-time changing driving environment.This allows unbiased and fair results based on flexible contextual understanding, even in situations that do not conform to rules and patterns, by considering the convergent interactions and dynamic situations of various objects that are possible in a real-time road environment. The transparent, environmentally adaptive, and fairness-based outcomes provide the basis for the decision-making process and support clear interpretation and explainability of decisions. All of these processes enable autonomous vehicles to draw reliable conclusions and take responsibility for their decisions in autonomous driving situations. Therefore, this paper proposes an Adaptability, Fairness, Transparency, Explainability and Accountability (AFTEA) Framework to build a stable and reliable autonomous driving environment in dynamic situations. This paper explains the definition, role, and necessity of AFTEA in artificial intelligence technology and highlights its value when applied and integrated into autonomous driving technology. The AFTEA framework with environmental adaptability will support the establishment of a sustainable autonomous driving environment in dynamic environments, and aims to provide a direction for establishing a stable and reliable AI system that adapts to various real-world scenarios

Keywords: contextual understanding in dynamic environments; situation awareness; trustworthy and sustainable AI; AFTEA

1. Introduction

The accelerating development of AI technology has revolutionized many fields, including healthcare, education, transport (autonomous driving), and finance [1], and is significantly improving the people's quality of life. As AI technology is capable of solving simple or complex problems, making decisions and automating more accurately and faster than humans, innovative AI-based technologies are spreading and being applied in various domains.

However, the complexity and diversity of the real world means that rule-based learning, which relies on specific data and scenarios, is limited in its ability to represent and reason about the convergent and diverse situations and scenes of the real world. In addition, AI technologies that focus on perfecting algorithm design and accuracy of conclusions are limited in their ability to flexibly adapt to various variables and characteristics of the real world, and are limited in their ability to produce results that are fair, adaptable, and transparently explain the causes and rationale behind the results, reflecting various social norms and rules shaped by humans. [2]

 doi:10.20944/preprints202408.0126.v1

Social phenomenon data should reflect various social phenomena, but if the training data does not reflect the situation of a particular group, or if it is trained with data that contains historical and socially unfair customs, disciplines, etc. This can lead to social disruption due to AI technologies that are not adapted to different social environments and are unable to solve social phenomena in different areas. If they are trained with a bias towards certain data or designed with algorithms that are not applicable to the problem they are trying to solve, they lack the ability to adapt to diverse data and cannot be used to solve complex social problems.From the perspective that AI algorithms should reflect complex social phenomena and situations, the algorithm structure is very complex, and it is difficult to intuitively interpret the input and output relationship that solves the problems of social phenomena.

Therefore, for the sustainable usage and improvement of AI algorithms, a clear rationale for the results must be explained. However, if the learning process of the data input to the algorithm and the causes and rationales for each result cannot be accurately visualized and explained, the clear rationale and causes of the results are ambiguous, leading to a black box-like transparency challenge.It should be transparent about what data the AI model is trained with, how it is trained in stages, how data contributes to each layer, and how predictions are made.

The problem with most AI models is that they have difficulty in clearly expressing the complex interrelationships between data, social issues, and data in the learning process, and in clearly visualizing and explaining the causes and rationale for the results obtained. XAI [3–5] is one of the concepts that has emerged to solve and complement this problem, and XAI aims to secure high reliability and stability by making the entire structure of decision-making explicable and interpretable.

However, XAI currently derives a verbal description of the visualized data or derives the contribution of each data characteristic applied to the learning process. In other words, in the case of explainable artificial intelligence (XAI) it is necessary to be able to derive the cause of each characteristic applied to the result and to explain what kind of predictive results will be produced by changing characteristics in real time, rather than a description limited to surface descriptions and probabilistic contributions. In addition, the social environment in which AI technology is applied may change in a different direction from the existing rules and methods, and many unexpected and sudden situations may occur. Conclusions and predictions based on existing rules and scenarios are difficult to reflect and respond to the changing social environment in real time, and there are limitations in deriving flexible results for the social environment. In a social environment where various laws, rules, and values are applied, AI needs to draw adaptive conclusions according to the actual field of application.

However, different laws, rules, and values are applied in each social environment where AI is applied, and it must be able to draw stable and flexible conclusions through contextual understanding and reasoning in dilemma situations that may violate laws, rules, etc. but require flexible judgment in the situation, and unexpected situations that deviate from existing laws and rules. AI technology can contribute to understanding social phenomena, overcoming and solving problems and difficulties that arise in society. However, laws, rules, and values vary widely in different social environments, and rules and customs in different contexts are intertwined, so data and learning that is limited to specific data and scenarios can lead to unfair judgements and biased decisions. It is therefore evident that if AI does not accurately reflect the various social structures and values present within a given social environment, it will be unable to reflect the diversity of that society and adapt and respond to social values and environments in a convergent manner. The rules and values that exist for the stability of the social environment may always be subject to dynamic changes and unexpected situations in a convergent society. In the absence of flexible understanding and reasoning about such situations, AI will be constrained in its ability to apply and respond to a range of situations and environments in real time. [6]

In order to overcome the limitations of AI and ensure the reliability and safety of AI, we are conducting research into reliable AI technology. This research is based on the conceptualisation and consideration of three key elements: Fairness, Accountability and Transparency(FAT). [7,8] FAT

 doi:10.20944/preprints202408.0126.v1

proposes the construction of AI models that consider fairness, accountability, and transparency as means of addressing the biases present in data and algorithms, the limited interpretability and explainability of models due to their complexity, and the lack of accountability for the issues caused by AI models. However, the concept of transparency in FAT entails that the implementation of the model's functioning can be interpreted in a transparent manner, thereby focusing solely on the derivation of results from the model. This model-centred approach aims to elucidate the inputs employed to generate the outputs at each neural network stage. However, it has limitations in terms of clearly delineating the data contributions and causal factors that led to the result. Consequently, following the FAT, the proposed FATE architecture incorporates explanations such as the interpretation of input data and interrelationships between data, as opposed to model-centered explanations.

In FATE(Explainability), [9–11] Explainability meaning the ability to interpret. In [], it emphasizes that the predicted outcome can be interpreted as a human would understand it. In [12], Both transparency and explainability are considered, so that the entire process of an AI algorithm is available for visualisation, explanation and interpretation. [9] explains that in most current AI systems, there is no clear explanation of how the AI produced a certain result or value, so being able to interpret and explain the resulting values is an essential element for advancing the application of AI in various fields.

In the real world, however, there are a number of factors at work, including a variety of objects, complex and unpredictable interrelationships between objects, and dynamic situations. The existing FAT and FATE models do not take environmental adaptation into account, which limits their ability to understand dynamic situations and respond to changes in a reliable manner. In the absence of the capacity to recognise novel situations or to derive intricate combinations of disparate data, it becomes challenging to regulate and respond in an effective manner to the perceived data. In order to ensure the stable application of AI in the real world, it is essential to derive and infer intricate relationships between objects, and to possess a clear understanding of novel situations and a rationale for appropriate responses to such situations. It is therefore necessary to analyze the various interrelationships between different situation perceptions and objects in order to gain an understanding of the new environment and situation. This allows the appropriate response and control methods for the situation to be inferred, thus ensuring a stable response and control in dynamic social environments.

In this paper, a new architectural framework is proposed, namely the Adaptability, Fairness, Transparency, Explainability and Accountability (AFTEA) architecture. This framework is designed to address convergent situations simultaneously. While numerous architectures and studies have considered the elements of an adaptive environment, either alone or in combination with some elements of FATE, relatively few have considered and utilized all five elements simultaneously.

AFTEA addresses the need to consider fairness, accountability, transparency, explainability, and environmental adaptability simultaneously and in combination across the entire process of contextualisation, understanding, reasoning, decision-making, and control. In order to support this necessity, this paper defines and offers an explanation of the manner in which the elements of AFTEA are defined and how they should be utilized in each process and stage of the entire AI process, from data collection to control.

In the next section, FAT and FATE are described, along with related work on how each element has been applied in AI. Then, in section 3, each element of AFTEA is defined and explained, and the AFTEA proposed in this paper is described in detail. Finally, section 4 summarizes and concludes the paper.

2. Related Works

2.1. Adaptability

In the interrelationship between autonomous vehicles (AVs) and human-driven vehicles (HVs), rule-based driving learning, which does not reflect the personal characteristics, personality, and

 doi:10.20944/preprints202408.0126.v1

preferences of human drivers, is challenging in predicting out-of-rule environments. [13]proposes an adaptive driving framework that combines human experience-driven driving with a rule base. The decentralized reinforcement learning framework optimizes experience-based utility and exposes learning agents to a wide range of driver exposure, becoming robust to human driver behavior and enabling cooperative and competitive control independent of the HV's aggression level and social preferences.Second, it prioritizes safety to avoid high-risk behaviors that could compromise driving safety. To this end, agent partial observability enables adaptive driving for moving objects with different characteristics in the environment.

[14] pointed out the difficulty of reasoning in a rule-centric paradigm and manually generated, limited rule-based autonomous driving technologies in responding to situations that require decisions that are out of sequence in the complexity and diversity of the real world. Therefore, [14] proposes a framework for constructing scenario semantic spaces in the human cognitive system to organize knowledge about new information and enable logical adaptation and interpretation.

2.2. Fairness

[15] investigated a number of real-world applications that were biased in different ways, examining the various biases and sources that can affect AI applications.This was followed by a categorization of fairness definitions, a taxonomy of fairness definitions that researchers have used to avoid existing biases in AI systems. It identifies different areas and subsections of AI for unfair consequences and describes how researchers have been trying to address them.

[16] categorized data bias and algorithmic bias as two potential reasons for unfairness in machine learning results. [16] categorized data bias into two types: distortions that distort what the machine learning algorithm learns and distortions that prevent it from making fair decisions based on the algorithms' operational behavior. In the context of decision-making in an autonomous environment, [16] proposed FairLight based on reinforcement learning, which emphasizes the importance of making decisions that are not biased towards any particular group. FairLight optimizes overall traffic performance by considering the fairness of the individual vehicles and enables fair and efficient traffic control through the allocation of efficient signal duration.

[17] presented the problem of bias in the training and distribution data of current AI models and the bias between distribution labels and sensitive groups. [17] introduces a correlation shift to account for fairness algorithms and group fairness. Second, [17] proposes a new preprocessing step. [17] samples the input data to reduce correlation shifts and formalize the data ratio adjustment between labels and sensitive groups to overcome the limitations of processing. This allows you to perform pre-processing correlation adjustments and unfairness mitigation on the processed data.

[18] proposed the concept of acting as an impartial judge in vehicle accidents and proposed an accident monitoring algorithm and accident investigation and analysis solution. This supports fair decision-making in autonomous vehicle accidents.

2.3. Transparency

[19] emphasizes that information that needs to be communicated to humans in autonomous vehicles must be transparently displayed to enable environmental understanding, and [19] emphasizes that Human Machine Interface (HMI) transparency for accurate and efficient communication is a critical factor in promoting safe driving. [19] evaluated five HMIs that integrated some or all of the following functions: information gathering, information analysis, decision making, and action execution. To verify the transparency principle, the transparency-based HMIs were evaluated for situational awareness, discomfort, and participant preference.

[20] presents a new problem of uncertainty by applying AI decision-making processes to autonomous driving. [20] argued that the autonomous driving environment comprises several interrelated components, including cybersecurity, robustness, fairness, transparency, privacy, and accountability, and therefore the various issues that arise in such a complex framework needs to be resolved.

 doi:10.20944/preprints202408.0126.v1

Accordingly, [20] presented an algorithm for the explainability of AV testing and validation in the Localization, Perception, Planning, Control, Human-Vehicle Interaction, and System Management phases of autonomous vehicles. By providing an understanding of the underlying algorithms and decision-making processes [20] ensures system safety and reliability, and evaluating performance in different scenarios, which promotes transparency and accountability.

2.4. Explainability

[21] proposed the need for AI to understand and reason about scene objects in the same way that humans solve problems. For scene understanding, only objects directly related to the driving task need attention and finite descriptions. In response, BDD-OIA is proposed to determine what drives an action and output relevant pairs of actions and descriptions to maximize explanatory power. BDD-OIA is a novel architecture for collaborative action/explanation prediction. It is implemented as an object detection module based on Faster RCNN and a global scene context module based on multi-task CNN. [21] utilized SG2VEC, a spatiotemporal scenario embedding methodology that uses graph neural networks (GNNs) and long short-range neural networks (LSTMs) to recognize visual scenes and predict future collisions through GNN and long short-term memory LSTM layers.

[22] proposes the need to describe AI-based decision-making processes that enable not only safe real-time decision-making in autonomous vehicles, but also compliance across multiple jurisdictions. We provide a thorough overview of current and emerging approaches for XAI-based autonomous driving. [22] presents a future direction and a new paradigm for XAI-based autonomous driving that can improve transparency, trustworthiness, and socialization by proposing an end-to-end approach to explanation.

[5] proposed a novel multimodal deep learning architecture to generate textual descriptions of driving scenarios that can be used as human-understandable explanations, collaboratively modeling the correlation between images (driving scenarios) and language (descriptions). [5] confirms that autonomous vehicles can effectively mimic the learning process of a human driver, and generating legitimate sentences for a given driving scenario enables appropriate driving decisions.

According to [23], AI technologies should explain and justify decisions in an understandable way to build trust and increase public acceptance. However, existing explanation techniques are primarily focus on explaining data-driven models (e.g., machine learning models) and are not well suited for complex goal-based systems like autonomous vehicles. Furthermore, these explanations are often only useful to experts and not easily utilized by the general user. [23] proposes an interpretable and user-oriented approach to explanation provision in autonomous driving with the goals of providing clarity and accountability. In an autonomous driving environment, a combination of object identification techniques, traffic conditions, scene graphs behavior, and road rules to generate descriptions. It focuses only on how this combination generates different descriptions (i.e., road rules) and suggests how different descriptions can be created and in which types of driving conditions descriptions are relevant.

2.5. Accountability

Two major challenges to overcome with autonomous vehicles are safety and liability. Society must hold autonomous vehicles accountable just as it holds humans behind the wheel accountable for their behavior and responsibility. One way to do this is to define terms and conditions for autonomous vehicles. [24] proposes five ways to make self-driving technology accountable: accepting responsibility by logging in before driving, configuring driving preferences to be accountable through scenarios, having some level of awareness of driving, inputting driving preferences, and a certification mark that says self-driving car manufacturers are accountable.

As described in section 2, there are components that are considered in AI technology and are being developed into robust AI technology. Each of these components has a powerful impact individually, and if each component is considered convergently, it is expected to lead to more advanced AI technology. In this case, it is essential to consider transparency and expalinability, which enable a

 doi:10.20944/preprints202408.0126.v1

transparent description of the situation and the providing of evidence-based criteria to derive clear validity and sustainable causality of the results. Based on transparent explanability, it is possible to understand results, provide prediction, inferences, decision-making, and control justification, and accountability is necessary for achieving reliable performance and results. When all of the elements of AFTEA interact convergently, robust control and predictable response is achieved. In the following section 3.1, each component is defined and described in detail, and the value of each component and the necessity of building AFTEA architecture that emphasises the convergence of each component is introduced.

3. AFTEA Architecture

AFTEA Architecture consists of Adaptabillity, Fairness, Transparency, Explainability, and Accountability as shown in Figure 1. In the real world, it is essential to collect data from many different and various objects and then to construct knowledge on the current state and situation in the environment. From the data collection stage, adaptability and fairness are at the most important considerations for clarity and contextualisation in the dynamic real world. It also considers fairness and adaptability at the stage of inferring new states and situations by generating knowledge that flexibly expresses the convergent real world and combining new data and existing knowledge.

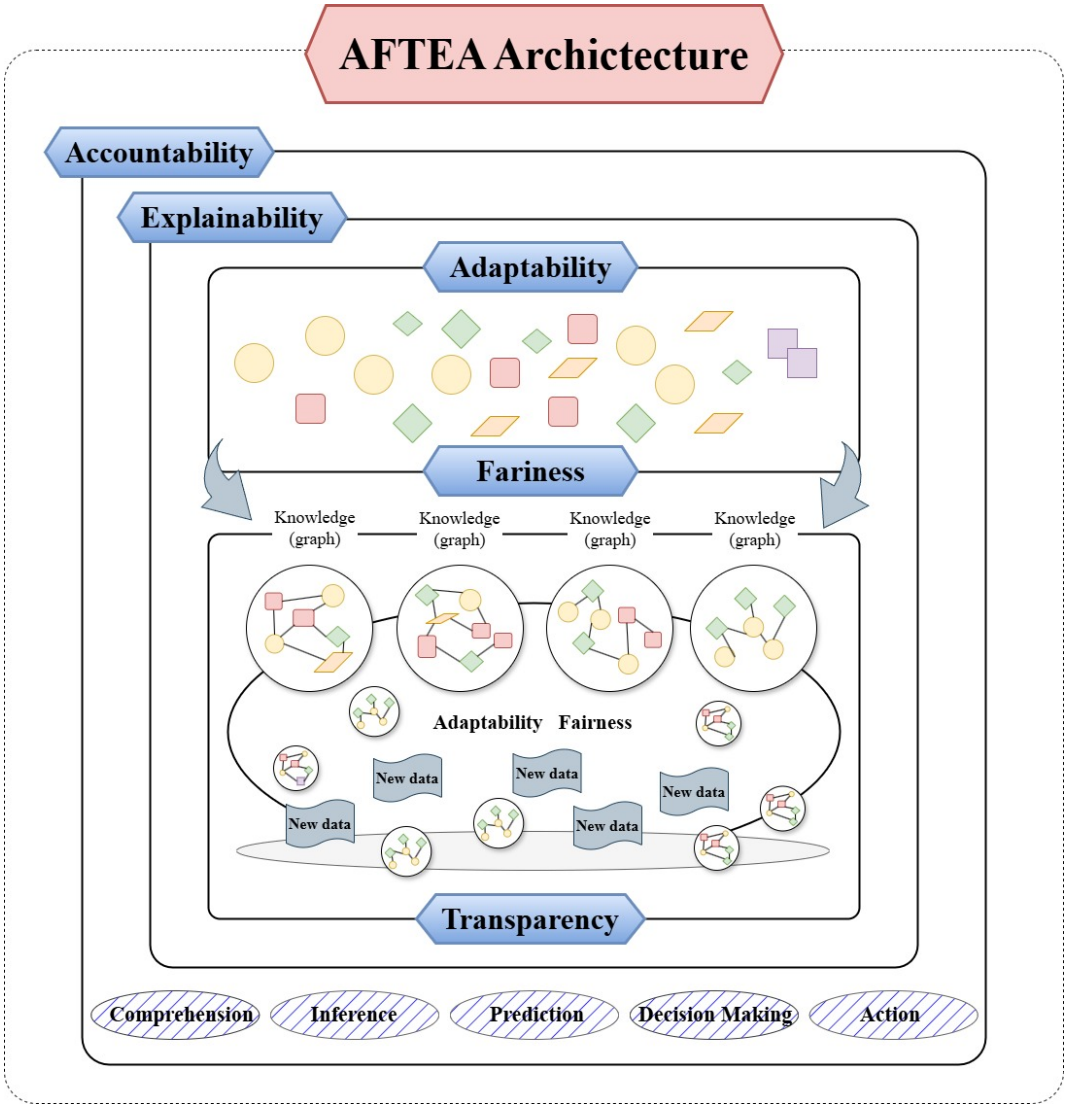

Figure 1. This is AFTEA Architecture.

 doi:10.20944/preprints202408.0126.v1

3.1. Adaptability

Adaptability is the ability to converge existing information from the real world with new information to anticipate and anticipate situations that may occur in different environments. Adaptive AI requires an inferential fusion of existing information and various information acquired in new situations to represent the dynamic real world and respond to various changes immediately.

For expressing such a state of affairs, the concepts of knowledge and ontology are being converged, and Graph Neural Network (GNN) [25–27] technology and Artificial General Intelligence (AGI) [28–30] technology are being researched to understand the complexity and diversity of the real world and to express convergent relationships and respond to immediate changes. Knowledge represents a situation by representing a given circumstance by storing complex structured data. Therefore, knowledge allows to derive relationships and rules between objects to infer a new situation from a known situation. To derive relationships and rules of objects, the concept of ontology is introduced, and ontology is applied to derive interactions and relationships of objects by interconnecting with knowledge. In the real world, various objects and information are interactively and convergently connected, so Knowledge graph technology [31–33] is applied to interconnect knowledge. Knowledge graph not only derives the relationship between objects by connecting them, but also expresses the relationship between the characteristics and meanings contained in the objects. Knowledge graph is developing into adaptive AI by representing the dynamic real world with the development of graph neural networks and enabling reasoning and response to various changes in situations. Therefore, adaptability is crucial for building knowledge from various objects and characteristics in the real world and inferring associations and causality between data in a dynamic environment that are changing in real time. In particular, in the context of real-world driving, a multitude of rules and patterns coexist, reflecting the diversity of social environments.

The interrelationships and combinations of objects within these environments give rise to a multitude of contexts and situations that deviate from the established rules and patterns. The driving environment requires adaptive decision-making not only for situations that occur in a regular pattern, but also for new situations that are irregular (outside the regular pattern) and different from previous experience.

It is therefore evident that in order to gain knowledge of the driving environment, it is necessary to develop the ability to construct a framework that incorporates the established rules and patterns of experience, and to derive knowledge that enables the inference and adaptation to new situations in real time, thereby facilitating an immediate response to the changing environment.

In order to generate adaptive conclusions to the situation by constructing a graph through knowledge, it is necessary to derive a flexible knowledge graph that accurately understands the existing knowledge graph and the new situation, allowing for complex convergence. It is extremely valuable to be able to infer and derive new knowledge graphs that can be compatible with existing knowledge graphs to infer possible actions in new situations and make the appropriate decisions. In other words, it needs to be adaptable to a variety of contextual information and multiple socio-environmental factors through flexible perception and interrelationships with unconventional objects.

Therefore, it is necessary to continuously establish and converge knowledge that enables the description of various situations, so that new patterns and rules arising from the knowledge graph of existing experiences become available for inference and appropriate response. Continuous and convergent knowledge graph construction allows for adaptive decision-making in various environments and contributes to establishing a safe autonomous driving environment.

The application of autonomous driving technology that is able to adapt to dynamic situations in real time is expandable by applying it to various domains that need to be applied to dynamic environments. In other words, in order to build adaptive algorithms that can be used in various social domains beyond the autonomous driving environment, it is essential to perform convergent contextualization, reasoning, and decision-making in various situations by consideration of the overall background and context of the current situation as well as deriving relationships between objects.

 doi:10.20944/preprints202408.0126.v1

3.2. Fairness

Fairness is a critical element in AI system-based decision-making processes to ensure that data and algorithms are unbiased and operate equitably across all scenarios and objects. The various environments in which AI systems are applied can be classified into dynamic and static environments. Static environments are those with minimal changes, where existing rule-based patterns can be applied without significant errors in the outcomes. However, in dynamic environments, such as autonomous driving, where objects move randomly and situations change in real-time, achieving stable training effects is challenging, and maintaining consistency in feature-based patterns is difficult. [34] Therefore, applying existing feature-based patterns in real-time dynamic situations is likely to inadequately consider environmental changes and the irregularity of moving objects, potentially leading to inappropriate situations that do not guarantee fairness. To ensure fairness, it is essential to carefully approach data collection and algorithm design so that AI systems operate impartially and equitably. [35]

First, to address fairness issues stemming from dataset characteristics, it is crucial to collect data from diverse sources to ensure representativeness and verify that the data is not biased towards specific situations or groups. [15,17] This helps in building a balanced dataset that includes a variety of scenarios and groups. In particular, for real-time data, efforts should be made to diversify and collect data from various time zones and locations to avoid concentration in specific times or places. For already collected datasets, it is necessary to check for overfitting caused by historical unfair practices reflected in the data, and to review and refine the labeling process to eliminate any bias. If these biased data characteristics are not considered in advance, it can negatively impact the model's performance and fairness.

Using biased data can cause the algorithm to produce results that are skewed towards specific categories, deviating from its original intended function and potentially failing to provide fair outcomes for all users. Efforts are made to minimize bias by training AI algorithms with the most fair datasets possible to avoid unfair results, but even after addressing data bias, the algorithm design process itself can still introduce unfairness. [16,36] Algorithm design focuses on problem-solving by extracting features of objects through comparative analysis of their histories and learning from these to predict unique situations. However, during the learning process to achieve results in a specific direction, the algorithm can become confined to a single domain-based scenario. This creates a risk that the objective function the algorithm aims to optimize may be biased under certain conditions.

If the specific features of the input data contain unbalanced or discriminatory information against certain groups, the model structure may react excessively sensitively to some features. This can lead to the objective function that the algorithm aims to optimize not ensuring overall fairness or acting unfavorably towards certain groups. Additionally, since existing rules and patterns extract and learn features based on past data, it is difficult to discover new variables for unseen elements and changed situations. These existing patterns often include biases, and following them as they are can undermine fairness.

In a truly dynamic environment, multiple objects are interrelated, and the algorithm, through comparative analysis of various features, may infer completely new types of information from numerous interactions, sometimes discovering incorrect correlations and making predictions based on them. For autonomous vehicles, where the environment in which the algorithm is trained can differ greatly from the environment in which it is applied, data collected from specific roads or conditions might not reflect other conditions. Consequently, the algorithm may make accurate predictions in some environments but malfunction in others. Such data bias and algorithm bias can be particularly pronounced in variable and unpredictable environments like autonomous driving. In other words, model fairness is an essential element for trustworthy AI. The process of building datasets and the functioning and operation of algorithms must be designed to be understandable and explainable. In addition, fair decision-making should be ensured by producing flexible and adaptive results that can merge with and adapt to new situations that deviate from existing rules and patterns, thereby avoiding biased outcomes. To achieve fairness, a situationally integrative approach is essential. This

 doi:10.20944/preprints202408.0126.v1

approach aims to produce unbiased outcomes by flexibly responding to new situations rather than being constrained by existing patterns. Through an integrative approach, it is essential to combine various factors and variables in diverse situations to understand the overall context and formulate appropriate strategies. A situationally integrative approach involves comprehensively analyzing the current situation and flexibly applying existing rules. This method is crucial for adapting to new situations and maintaining fairness.

Therefore, by ensuring fairness through three approaches: data bias, algorithmic bias, and environmental adaptive bias, AI in autonomous driving environments can produce unbiased results in all conditions, deliver accurate and reliable outcomes, and operate fairly for diverse users. As a result, these approaches enhance the transparency and reliability of AI systems, helping them operate fairly for diverse user groups. This ensures that AI systems can adapt to dynamically changing real-time environments and respond effectively to various situations. Ultimately, AI systems that ensure fairness can build social trust and contribute to the increased utilization and acceptance of AI technology.

3.3. Accountability

Accountability is the capacity to judge the justification of actions for decisions and adjust behavior according to situations. The accountability of artificial intelligence technology requires that when conclusions are generated by applying artificial intelligence technology, there must be justification for the generated conclusions, and it must be able to be safely and reliably controlled in various situations. Accountability for AI outcomes requires transparency and accountability for learning and reasoning that leads to unbiased and fair outcomes, and transparency and accountability for clearly explaining the reason and evidence behind results.

It is crucial to maintain responsibility for achieving stable results and making effective decisions in every circumstance, regardless of prior exposure. For the responsible use of AI, it is imperative to substantiate how effectively the factors of fairness, adaptability to environmental contexts, transparency, and explainability are incorporated across the entire AI algorithm process. From a societal perspective, it is necessary to ensure that legal regulations and rules are correctly and validly applied to the domains in which AI algorithms are applied, and that they are generally applicable by providing reliable results to the users of the applications.Therefore, accountability in AI technology must consider both technical and social accountability, and it becomes possible to apply responsible AI technology that can be applied to society when both are combined.

The autonomous driving environment where AFTEA is focused requires accountability for safe outcomes since human lives are involved. Unlike environments that rely on rule-based, learned situations, autonomous environments encounter many dynamic situations, including situations that are out of line with the rules, situations that are different from existing learned situations, and unexpected situations. It is necessary to be responsible for ensuring that fair conclusions are achieved that are adapted to a variety of changing factors, such as the surrounding environment and road driving regulations. Also, AI processes require accountability for transparent rationale and cause-based explanations, and accountability for ensuring that decisions are implemented reliably to enable stable driving control. Fairness in AI requires accountability for bias in data and models. To eliminate bias in a dataset, it is important to be able to assess whether the representation of a group of datasets is fair, and to determine whether the data in the group is unbalanced and whether it is the group of datasets from which biased data values are derived. If datasets contain data labels, it is also important to determine if there is a bias in the names of the data labels and the distribution of the data labels.

For accountability against model bias, it is necessary to evaluate the extent to which the model's predicted probabilities are consistent with actual consequences. To ensure a sustainable model, it should be possible to determine how well predicted outcomes match actual outcomes, and to derive a clear error rate for the predicted values, and a process for refinement.

Transparency and explainability accountability requires transparent indicators of how the results were derived, what the results are, and what factors led to the results and the rationale for the results.

 doi:10.20944/preprints202408.0126.v1

Accountability should be based on generalizable metrics, such as verification of the accuracy of the model and results descriptions from an explanatory perspective, and validation of the rationale for the conclusions against existing knowledge and expert domains. Accountability in environmental adaptability should be assessed by evaluating the ability to effectively learn from new environments and new data, and to fuse old and new data to produce reliable results and decisions in a variety of environments.

To ensure accountability for AI's environmental adaptability, it can be divided into the aspects of robustness to changes in data and models, and robustness to changes in the situation and environment in which AI technology is applied. In the case of data, it is necessary to be responsible for how robust and stable it is in the face of various changes in input data, noise, and attacks. If the reliability and robustness of the model is guaranteed when the same patterns and features of the input data are input during the learning process, it will result in highly accurate results for the input data.

However, in the real world, where there are various environmental changes, data with subtle changes and patterns and characteristics that do not exactly match the data utilized for training are input, so it is essential to be able to accurately reflect these data changes and produce results that are appropriate for the changed data. It should be able to understand new environments and reason about outcomes based on existing learned results, while simultaneously assessing whether it can adapt reliably to situations that are rare and non-common scenarios. Accountability is not just about the consequences of the results produced.

It must be able to provide a clear justification for learning, understanding, reasoning, situation awareness, prediction and decision-making, and control that takes into account all the elements of fairness, transparency, explainability, and environmental adaptability that AI technology requires. Accountability in AI requires the ability to give reliable validity to the results obtained, and to judge the rightness or wrongness of actions and controls when making decisions, so that clear criteria for control and response can be established for the domain in which the AI technology is applied. Also, Accountability has to be achieved by using AI technology to combine elements of AFTEA by providing valid assessments and criteria to draw conclusions that ensure the robustness and reliability of all components of AFTEA.

3.4. Transparency

Transparency is a crucial concept in many fields, signifying that information and processes are clear and publicly accessible. It is an important factor in enhancing reliability and fairness, and it enables the explanation and verification of results. Transparency can generally be divided into 'Information and Algorithmic Transparency', 'Transparency in Result Derivation and Decision-Making Processes', and 'Transparency for Accountability.'

First, 'Information and Algorithmic Transparency' means that datasets and algorithms should be publicly available. By making the algorithms transparent about which datasets the system analyzed and learned from to make informed decisions, you build trust in the results it produces. This allows us to evaluate the fairness of the system and trace the source of errors if they occur. Understanding the relationship between data and algorithms and explaining the decision-making process of machine learning models is important and can contribute to increasing the reliability of AI systems and promoting the development of socially transparent technologies. [37,38]

Second, 'Transparency in Result Derivation and Decision-Making Processes' means clearly showing how results are generated and transparently presenting the outcomes. Especially in autonomous driving systems, In autonomous driving systems, it is essential to explain why the vehicle chose a particular action. [39] This provides information about the system's state and helps users understand and trust the automated system's operating principles and decision-making rationale. For example, if an autonomous car suddenly slows down, it should be able to clearly explain whether the reason was an obstacle on the road or the movement of another vehicle. Through these principles of transparency, it is essential to clearly define and provide the necessary information so that users can understand

 doi:10.20944/preprints202408.0126.v1

the system's intentions. [19,40] In other words, clearly explaining the reasons behind the system's decisions is crucial for enhancing human trust. [41]

Liu et al. (2022) [40] proposed a functional transparency(FT) assessment approach to address the limitations of existing Human Machine Interface(HMI) transparency evaluation methods that rely on the quantity of information. Unlike traditional transparency, which merely emphasizes the amount of information provided, functional transparency (FT) focuses on how well the HMI can be understood by the user after interaction. This approach evaluates how effectively the HMI design enables users to understand the environment based on the information transmitted, and it reexamines the effectiveness and importance of the information delivery methods.

Polam et al. (2019) [19] aim to extract the information needed by drivers in the design of HMI for autonomous vehicles, helping drivers to understand and trust the behavior of the autonomous driving system. This study considers Driver-Vehicle-Environment (DVE) conditions and driver status, using a rule-based algorithm to visually clarify why an autonomous vehicle is reducing its speed, thereby aiding drivers in understanding the system's intentions. This approach enhances drivers' situational awareness (SA) and improves the transparency of the system.

Thirdly, 'Transparency for Accountability' links the results obtained and the supporting evidence to the system's accountability by transparently disclosing them. This means that newly generated and accumulated data, derived results, and interpretations of decision-making must be openly and transparently available in real-time. In the event of an accident in an autonomous system, this information should be transparently disclosed in order to reconstruct the chronological sequence of events to analyze the cause of the accident and interpret responsibility. [42] Through this, it should be clearly identified on what data decisions were based and how those decisions were made, to clearly determine responsibility and derive improvement measures for similar situations in the future.

The event data recorder(EDR) in an autonomous vehicle records data related to the operation of the vehicle. These data can reconstruct the events leading up to an accident and provide important information for legal proceedings and insurance claims. Researchers are developing algorithms to more accurately analyze post-accident data. [43]

JN Njoku et al. (2023) [42] propose an innovative concept that uses recorded data and location-based identification to ensure fair judgment in vehicle accidents. Their research demonstrates the feasibility of the proposed solution for accident investigation and analysis.

A Rizaldi et al. (2019) [24] addresses the problem of ensuring that autonomous vehicles follow traffic rules and clarify responsibility in the event of a collision. To solve this problem, they propose a method to make traffic rules datafied and mechanically testable. They show that if traffic rules are precise and unambiguous, vehicles can avoid collisions while obeying traffic rules, which is important for establishing liability. This contributes that the behavior of autonomous vehicles is transparently evaluated and that responsibility is clearly identified.

D Omeiza et al. (2021) [23] propose an interpretable tree-based user-centered approach to describe autonomous driving behavior. One way to ensure multiple accountability is to provide a description of what the vehicle 'saw', did, and can do in a given scenario. To this end, based on hazard object identification in driving scenes and traffic object representation using scene graphs, we combine observations, autonomous vehicle behavior, and road rules to provide interpretable tree-based descriptions. A user study evaluating the types of explanations in different driving scenarios emphasizes the importance of causal explanations, especially in safety-critical scenarios.

In this way, the data, results, and decision-making processes related to autonomous vehicles must be formalized and transparently disclosed so that responsibility can be clearly identified and improvements can be made in similar situations. A lack of transparency can lead to a variety of problems, including a lack of trust, questions about fairness, and difficulties in accurately analyzing system errors. In autonomous driving systems in particular, a lack of transparency can significantly undermine the trust of users and the general public. Since autonomous driving systems are critical

 doi:10.20944/preprints202408.0126.v1

systems that directly impact human lives, ensuring safety and reliability through transparency is essential.

Therefore, autonomous driving systems need to ensure transparency through the disclosure of datasets and algorithms, clear explanations of how results are derived, and transparent interpretations of results and decisions. This enhances the explainability of outcomes and helps provide reliable decision-making. Transparency is a crucial element that provides clarity and legitimacy to the results, playing a key role in ensuring the safety and reliability of autonomous driving systems.

3.5. Explainability

Explainability is the interpretation of the basis and causes of the results obtained to demonstrate the validity and clarity of the results obtained. The result of AI is a complex inference value of the entire process from the data recognition step in the situations and environments to be recognized to the interaction between the perceived object data and the situation awareness to the result. By interpreting the state, phenomenon, and situation of each process from the cognitive stage to the final decision, the explanation of the cause of the situation can improve the clarity of convergent reasoning, and reliable explanations are achieved through transparent evidence and accurate information.Reliable explanation-based learning-based information and knowledge generation enables sophisticated and flexible decision-making for previously learned situations. In unexperienced situations, it enables convergent intelligent reasoning and understanding to generate knowledge and information that are applicable to a variety of situations. Therefore, the following process should allow for flexibility in deriving the explanation in each process.

Explainability at the recognition stage should be able to identify the recognition results by visually deriving the elements of each recognised object from the recognition of data in the model. The visual representation and derivation of recognition results enables status tracking and continuous monitoring, and enables clear causes and rationales for the results when notifying and controlling the AI about the situation. Providing a reasonable basis and causal factors for object recognition provides a clear visual explanation of the factors that contribute to situations, abnormal situations, etc.

The real world is composed of convergent interactions of objects, but even though each individual object can be perceived and described, it is difficult to perceive and understand the situations that they constitute if their relationships are not deduced and described. Deriving the Interrelation of objects can derive the grounds and causes for the perceived objects and situations. By explaining these grounds and causes, it is possible to infer the relationships between objects formed in various and new situations and to continue to provide valid grounds and causes for the perception of new situations. It can derive the priority and importance of objects for the formation of mutual relationships among objects and provide the basis for forming mutual relationships.

Knowledge is needed to represent and reason about interactions. Knowledge is represented in AI as a knowledge representation, which is a way of representing contextual information by describing situations that occur in the real world. The knowledge representation becomes the basis for making decisions and performing control over the situation by deriving the perceived situation information. The relationships between objects formed in consideration of interaction construct various knowledge for judging the situation. When a new situation is recognised by forming knowledge, it is possible to form new knowledge for decision-making and control through a reasoning process suitable for unlearned situations by fusing the interrelationships of objects and existing knowledge, and to present decision-making and control criteria for that situation.

The knowledge formed by the convergent interrelationships between objects becomes the basis for decision-making and control in various situations, and the process of knowledge formation, deriving labels and descriptions of the knowledge on what basis the derived knowledge provides for decision-making, is essential for reliable decision-making. By linking the above interactions to the process of performing them, it should be possible to clearly explain how the interactions between the objects

 doi:10.20944/preprints202408.0126.v1

contained in the knowledge were applied, which objects and what weights were derived to build the knowledge appropriate to the situation.

In the case of new situation recognition, it is possible to express and recognise the situation by explaining the complex reasoning process on which of the existing knowledge should be selected and how it fuses with the Interrelation of new objects, and to derive detailed criteria and specificity for situation-specific decision-making and control. When a new situation occurs, the Interrelation of the perceived objects enables an applied understanding of the situation through convergence of existing knowledge, and new knowledge is constructed by combining new Interrelation with existing knowledge, so it is necessary to explain the selection criteria for valid knowledge selection.

When recognising a situation, it is necessary to present the basis and criteria for whether the Interrelation of objects can select knowledge based on the learned situation. If there is no learned situation and new knowledge needs to be built with new situation cognition, the criteria of the selection range should be presented to see what existing knowledge can be utilised through the Interrelation of objects, and if so, the explanation of which part and how to utilise it should be presented. Then, in the process of composite fusion of new interrelationships between objects and existing knowledge, the explanation of each part of how the new relation is connected and fused with the existing knowledge can be presented to improve the validity of composite reasoning for building new knowledge.

Various situations in the real world do not always occur in a certain pattern due to the diversity of perceived data and various interactions. Explainability in AI should not only explain the basis and cause of conclusions, but also derive valid factors for prediction, decision making, judgement, and control from the process of data recognition and cognition to the process of situation recognition, understanding, and reasoning. Explainability through clear evidence and factors for the results derived from each process enables more diverse information reasoning and interpretation with detailed evidence and cause interpretation, and enables flexible situation-specific response and decision making.

4. Conclusion

This paper proposed the AFTEA Framework to define the essential factors to be regarded in the learning, reasoning, and decision-making process of artificial intelligence, and suggested the direction for the improvement of stable and reliable autonomous driving environment and sustainable artificial intelligence technology. The current AI technology emphasizes the need for accountability, fairness, transparency, and explainability (or ethics). However, in the case of AI technology, it is imperative to consider contextual and environmental adaptive factors as it must be compatible with a highly dynamic social environment such as the autonomous driving environment and real world, where diverse changes take place in real time.

However, much of the current research is based on FAT and FATE, and there is a lack of research on architectures that converge the previous elements, including adaptive elements. Therefore, this paper proposes an Adaptability, Fairness, Transparency, Explainability and Accountability (AFTEA) architecture to ensure a stable, clearly informed, and contextually adaptable approach to the dynamic real world.

This paper explains the need for AFTEA by defining the overall architecture of AFTEA, each of its components, and describing how each component is applied in the process of performing AI.

Adaptability is the ability to fuse existing information and new knowledge in the real world. It is a characteristic that enables immediate response through inferential fusion of existing information and new situations by predicting and reasoning about situations that occur in various environments.

Fairness is essential for trustworthy AI, making it possible to understand and explain how datasets are built and how algorithms function and work. It enables contextualization, context-adaptive and flexible outcomes, and fair decision-making in new situations that challenge established patterns and deviate from the rules.

Transparency refers to the clarity and public accessibility of information and processes and is an essential element for increasing reliability and fairness and enabling explanation and verification

 doi:10.20944/preprints202408.0126.v1

of results. Transparency refers to the clarity and public accessibility of information and processes and is an essential element for reliability and fairness and enabling explanation and verification of results. Accordingly, this paper specifies the need for transparency in AI by dividing transparency into transparency of information and algorithms in general, transparency of results and decision-making processes, and transparency for accountability.

In the last place, explainability is the interpretation of the reason and cause of the derived results to verify the validity and clarity of the results. As it is a complex inference value of the entire process from data collection to data recognition in the situation and environment, interaction between recognized objects and data, and situation awareness, it enables rational decision-making by presenting not only the basis for the final result but also a detailed explanation of the entire process of deriving the result.

Accountability is the ability to judge the legitimacy of decisions and control actions according to the situation, and is a factor that enables the justification of conclusions drawn by AI and the safe and reliable control of decisions in various situations.

As described above, this paper described and defined each element of AFTEA and proposed an architecture that can be convergently applied to the real world, including environmental adaptability. By including the environmental adaptability element, AFTEA provides a direction to improve from the existing algorithms that consider fairness, accountability, explanability, and transparency to algorithms that analyze situations that occur in various environments and are able to adapt and respond in real time. Based on the AFTEA architecture, we will extend the research to experiment and apply the architecture in a real autonomous driving environment. This research will be conducted to derive numerical and visual validation of the AFTEA architecture in real-world scenarios. In addition, AFTEA architecture will be enhanced to become a sustainable AI technology in various real-world domains beyond the autonomous driving environment.

Author Contributions: SB and JE conceptualized ideas and the framework, investigated related works, developed the theory, created and designed the architecture and wrote the main manuscript text and figures. YI supervised the completion of the work, contributed to manuscript preparation, funded acquisition and administered project. All authors reviewed the manuscript. All authors read and approved the final manuscript.

Funding: This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) program(IITP-2024-RS-2022-00156299) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation) and supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT). (No. NRF-2023R1A2C1005779).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Connor, S.; Li, T.; Roberts, R.; Thakkar, S.; Liu, Z.; Tong, W. Adaptability of AI for safety evaluation in regulatory science: A case study of drug-induced liver injury. *Frontiers in Artificial Intelligence* **2022**, *5*, 1034631.
2. Liu, H.; Wang, Y.; Fan, W.; Liu, X.; Li, Y.; Jain, S.; Liu, Y.; Jain, A.; Tang, J. Trustworthy ai: A computational perspective. *ACM Transactions on Intelligent Systems and Technology* **2022**, *14*, 1–59.
3. Rane, N.; Choudhary, S.; Rane, J. Explainable Artificial Intelligence (XAI) approaches for transparency and accountability in financial decision-making. *Available at SSRN 4640316* **2023**.
4. Alikhademi, K.; Richardson, B.; Drobina, E.; Gilbert, J. Can explainable AI explain unfairness? A framework for evaluating explainable AI. arXiv 2021. *arXiv preprint arXiv:2106.07483*.
5. Omeiza, D.; Webb, H.; Jirotka, M.; Kunze, L. Explanations in autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems* **2021**, *23*, 10142–10162.
6. Novelli, C.; Taddeo, M.; Floridi, L. Accountability in artificial intelligence: what it is and how it works. *AI & SOCIETY* **2023**, pp. 1–12.

7. Shin, D.; Park, Y.J. Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior* **2019**, *98*, 277–284.
8. Lepri, B.; Oliver, N.; Letouzé, E.; Pentland, A.; Vinck, P. Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges. *Philosophy & Technology* **2018**, *31*, 611–627.
9. Shin, D. User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. *Journal of Broadcasting & Electronic Media* **2020**, *64*, 541–565.
10. Quttainah, M.; Mishra, V.; Madakam, S.; Lurie, Y.; Mark, S.; et al. Cost, Usability, Credibility, Fairness, Accountability, Transparency, and Explainability Framework for Safe and Effective Large Language Models in Medical Education: Narrative Review and Qualitative Study. *JMIR AI* **2024**, *3*, e51834.
11. Shaban-Nejad, A.; Michalowski, M.; Brownstein, J.S.; Buckeridge, D.L. Guest editorial explainable AI: towards fairness, accountability, transparency and trust in healthcare. *IEEE Journal of Biomedical and Health Informatics* **2021**, *25*, 2374–2375.
12. Diakopoulos, N.; Koliska, M. Algorithmic transparency in the news media. *Digital journalism* **2017**, *5*, 809–828.
13. Valiente, R.; Toghi, B.; Pedarsani, R.; Fallah, Y.P. Robustness and adaptability of reinforcement learning-based cooperative autonomous driving in mixed-autonomy traffic. *IEEE Open Journal of Intelligent Transportation Systems* **2022**, *3*, 397–410.
14. Li, X.; Bai, Y.; Cai, P.; Wen, L.; Fu, D.; Zhang, B.; Yang, X.; Cai, X.; Ma, T.; Guo, J.; et al. Towards knowledge-driven autonomous driving. *arXiv preprint arXiv:2312.04316* **2023**.
15. Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; Galstyan, A. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)* **2021**, *54*, 1–35.
16. Ye, Y.; Ding, J.; Wang, T.; Zhou, J.; Wei, X.; Chen, M. Fairlight: Fairness-aware autonomous traffic signal control with hierarchical action space. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **2022**, *42*, 2434–2446.
17. Roh, Y.; Lee, K.; Whang, S.E.; Suh, C. Improving fair training under correlation shifts. In Proceedings of the International Conference on Machine Learning. PMLR, 2023, pp. 29179–29209.
18. Njoku, J.N.; Nwakanma, C.I.; Lee, J.M.; Kim, D.S. Enhancing Security and Accountability in Autonomous Vehicles through Robust Speaker Identification and Blockchain-Based Event Recording. *Electronics* **2023**, *12*. https://doi.org/10.3390/electronics12244998.
19. Pokam, R.; Debernard, S.; Chauvin, C.; Langlois, S. Principles of transparency for autonomous vehicles: first results of an experiment with an augmented reality human–machine interface. *Cognition, Technology & Work* **2019**, *21*, 643–656.
20. Llorca, D.F.; Hamon, R.; Junklewitz, H.; Grosse, K.; Kunze, L.; Seiniger, P.; Swaim, R.; Reed, N.; Alahi, A.; Gómez, E.; et al. Testing autonomous vehicles and AI: perspectives and challenges from cybersecurity, transparency, robustness and fairness. *arXiv preprint arXiv:2403.14641* **2024**.
21. Xu, Y.; Yang, X.; Gong, L.; Lin, H.C.; Wu, T.Y.; Li, Y.; Vasconcelos, N. Explainable object-induced action decision for autonomous vehicles. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9523–9532.
22. Atakishiyev, S.; Salameh, M.; Yao, H.; Goebel, R. Explainable artificial intelligence for autonomous driving: A comprehensive overview and field guide for future research directions. *IEEE Access* **2024**.
23. Omeiza, D.; Web, H.; Jirotka, M.; Kunze, L. Towards accountability: Providing intelligible explanations in autonomous driving. In Proceedings of the 2021 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2021, pp. 231–237.
24. Rizaldi, A.; Althoff, M. Formalising traffic rules for accountability of autonomous vehicles. In Proceedings of the 2015 IEEE 18th international conference on intelligent transportation systems. IEEE, 2015, pp. 1658–1665.
25. Sadid, H.; Antoniou, C. Dynamic Spatio-temporal Graph Neural Network for Surrounding-aware Trajectory Prediction of Autonomous Vehicles. *IEEE Transactions on Intelligent Vehicles* **2024**.
26. Bi, W.; Cheng, X.; Xu, B.; Sun, X.; Xu, L.; Shen, H. Bridged-gnn: Knowledge bridge learning for effective knowledge transfer. In Proceedings of the Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, 2023, pp. 99–109.
27. Ye, Z.; Kumar, Y.J.; Sing, G.O.; Song, F.; Wang, J. A Comprehensive Survey of Graph Neural Networks for Knowledge Graphs. *IEEE Access* **2022**, *10*, 75729–75741. https://doi.org/10.1109/ACCESS.2022.3191784.

 doi:10.20944/preprints202408.0126.v1

28. Goertzel, B. Artificial general intelligence: concept, state of the art, and future prospects. *Journal of Artificial General Intelligence* **2014**, *5*, 1.
29. Goertzel, B.; Pennachin, C. *Artificial general intelligence*; Vol. 2, Springer, 2007.
30. Baum, S. A survey of artificial general intelligence projects for ethics, risk, and policy. *Global Catastrophic Risk Institute Working Paper* **2017**, pp. 17–1.
31. Zhang, J.; Zan, H.; Wu, S.; Zhang, K.; Huo, J. Adaptive Graph Neural Network with Incremental Learning Mechanism for Knowledge Graph Reasoning. *Electronics* **2024**, *13*, 2778.
32. Feng, S.; Zhou, C.; Liu, Q.; Ji, X.; Huang, M. Temporal Knowledge Graph Reasoning Based on Entity Relationship Similarity Perception. *Electronics* **2024**, *13*, 2417.
33. Li, Y.; Lei, Y.; Yan, Y.; Yin, C.; Zhang, J. Design and Development of Knowledge Graph for Industrial Chain Based on Deep Learning. *Electronics* **2024**, *13*, 1539.
34. Lei, X.; Zhang, Z.; Dong, P. Dynamic path planning of unknown environment based on deep reinforcement learning. *Journal of Robotics* **2018**, *2018*, 5781591.
35. Bird, S.; Dudík, M.; Edgar, R.; Horn, B.; Lutz, R.; Milan, V.; Sameki, M.; Wallach, H.; Walker, K. Fairlearn: A toolkit for assessing and improving fairness in AI. *Microsoft, Tech. Rep. MSR-TR-2020-32* **2020**.
36. Danks, D.; London, A.J. Algorithmic Bias in Autonomous Systems. In Proceedings of the Ijcai, 2017, Vol. 17, pp. 4691–4697.
37. Larsson, S.; Heintz, F. Transparency in artificial intelligence. *Internet policy review* **2020**, *9*, 1–16.
38. Kemper, J.; Kolkman, D. Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society* **2019**, *22*, 2081–2096.
39. Oliveira, L.; Burns, C.; Luton, J.; Iyer, S.; Birrell, S. The influence of system transparency on trust: Evaluating interfaces in a highly automated vehicle. *Transportation research part F: traffic psychology and behaviour* **2020**, *72*, 280–296.
40. Liu, Y.C.; Figalová, N.; Bengler, K. Transparency assessment on level 2 automated vehicle hmis. *Information* **2022**, *13*, 489.
41. Cysneiros, L.M.; Raffi, M.; do Prado Leite, J.C.S. Software transparency as a key requirement for self-driving cars. In Proceedings of the 2018 IEEE 26th international requirements engineering conference (RE). IEEE, 2018, pp. 382–387.
42. Njoku, J.N.; Nwakanma, C.I.; Lee, J.M.; Kim, D.S. Enhancing Security and Accountability in Autonomous Vehicles through Robust Speaker Identification and Blockchain-Based Event Recording. *Electronics* **2023**, *12*, 4998.
43. Kropka, C. *"Cruise"ing for "Waymo" Lawsuits: Liability in Autonomous Vehicle Crashes*; Richmond: Richmond, VA, USA, 2016.