ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940www.ijasem.org

Vol 9, Issuse.2 April 2021

Smart city security, data management, and ethical issues will be examined in the development of future human-centered smart cities.

B.Venkateswarlu[1],Dr.K.G.S.Venkatesan[2],Dr.K.Venkata Ramana[3],S.Sreehari Raju [4],

Abstract

There is a pressing need to consider the long-term viability of livable cities in light of the world's growing population. While artificial intelligence services are becoming more common in modern smart cities, it's important to remember that (1) technology can facilitate prosperity, well-being, urban livability or social justice only when it's paired with the right analogue complements (such as well-thought-out policies, mature institutions, responsible governance) and (2) the ultimate goal of these smart cities is to facilitate and enhance human welfare and social flourishing. Extremism, polarisation, disinformation, and Internet addiction have all been linked to a variety of technical business models and characteristics, according to recent research. These findings highlight the critical relevance of resolving philosophical and ethical issues related to the development of AI algorithms that will serve as the foundation for future cities' technological infrastructure. There are requests for technology to be more humanitarian and human-centered across the world. Key barriers to a successful deployment of AI in human-centric applications are examined and explored in this study with a special focus on the convergence of these concepts/concerns, which include security, robustness, interpretability, and ethical (data and algorithmic) challenges. These important issues are examined in depth, and we examine how one of these challenges may lead to or help solve otherchallenges. Research in these areas is also advised on how to fill up the present gaps and find better answers, and this publication does just that. In our opinion, this level of thoroughness is essential for future study in the field.

Introduction

By 2050, it is predicted that 66% of the world's population will live in cities, a figure that now stands at 54%. Rapid urbanisation is spurred by economic factors, but the environmental and social costs are important as well. In order to maintain a balance between rising urbanisation and the limited resources available in cities, environmental, social, and economic sustainability are essential. As a result of current technology, we are attempting to enhance the environmental and economical elements of urban living, and to mitigate the accompanying issues. Modern technologies such as Internet of Things (IoT) sensors are being used to gather and analyse data on many aspects of urban life in smartcities [2,3]. [2,3]. People from several fields, such as engineering, architecture, urban planning, and economics, must work together to develop, design, build, and deploy a smart solution for a specific purpose. A variety of IoT sensor data may be analysed using AI approaches to better manage and make use of available resources via the use of artificial intelligence (AI). Data science, statistical learning, machine learning, and deep learning are all examples of AI used broadly in this paper as an um-brella term for techniques and algorithms that can learn from data. Intelligent systems that can perform tasks such as perception, reasoning, and inference are also included (i.e., expert systems,probabilistic graphical models, Bayesian networks). Artificial Intelligence (AI) may help you get insights from data if you know how to ask the correct questions and are aware of the hazards, according to Greg Stone [4]. AI has shown to be very useful in a wide range of smart city applications, including healthcare, transportation, education, the environment, agriculture, and military [5–9]. This study focuses on artificial intelligence (AI) technology for smart cities, however our concepts can be applied to a broader range of AI technology for smart cities. The most recent substantial advancements have also been made feasible by improvements in artificial intelligence. We will often use the terms "AI safety" and "AI ethics" interchangeably because of their close connection to AI.

Professor[1,2,3,4,] Assistant Professor[1,2,3,4] ,Associate Professor[1,2,3,4],
Department of CSE Engineering,
Pallavi Engineering College,
Mail ID:bvenkat1109@gmail.com, Mail ID:venkatesh.kgs@gmail.com,
Kuntloor(V),Hayathnagar(M),Hyderabad,R.R.Dist.-501505.

Data from individual sensors or the combined data of several sensors may be processed using AI algorithms, which can then give helpful information on how to improve underlying services in a smart city. Data from various sections of a city (such as roads, commuting modes, and passenger counts) may be analysed by AI in the context of transportation for the sake of future planning and deployment of various transportation systems inside a city. Many dangers and problems must be overcome before AI indifferent smart city applications can be effectively deployed [10–12]. According to [13,14], algorithmic predictions may be prejudiced towards specific races and genders, which might have a negative impact on sensitive human-centric applications. Smartcities' AI applications face other dangers than data-centric ones. For example, adversarial assaults on AI models may be launched in a variety of ways to damage the models' predictive2capabilities. Sophisticated applications, such as driverless cars, are vulnerable to assaults that may have a devastating impact on human lives and infrastructure [15]. The lack of interpretability (i.e., humans' inability to grasp the source of an AI model's choice) [16] is another major obstacle to the use of AI in smart cities. In important smart city applications, the predictive skills of AI models are not sufficient to address a problem entirely; rather, the reasons behind the forecast are required to be understood[17,18]. Amazon's AI recruiting tool was found to be biassed against women [19], and the company's Face Rekognition, a gender recognition tool was found to be 31.4% less accurate in classifying the gender of dark-skinned men than light-skinned men [20,21]. It also helps to ensure that AI decisions in an under-lying application are equitable by avoiding decisions based on protected attributes (e.g., race, gender, and age), and ensuring Concerns about the use of AI algorithms in human-centric smart city applications have grown steadily in recent years [20,23,24], for example, to ensure privacy problems in surveillance systems, uneven inclusion of residents in various services, and biases in predictive policing. There may be a ripple effect if one of these difficulties is solved. For example, explainable may be used to counteract the issue of choice interpretability and bias. The explanations supplied by explainable AI, on the other hand, may enable the attackers design more unfavourable assaults. AI may also help protect against adversarial attacks.

AI-based smart city applications

The sensors in a smart city collect data on diverse elements of the city (e.g., transportation, healthcare, and the environment), which is then transferred to a central server for analysis or processed locally at the edge devices to get relevant insights using artificial intelligence (AI) methods. Government authorities may now collect real-time data, along with the capabilities of AI, in order to better administer public services in cities. Cities may minimise traffic, crowds, and pollution by removing bottlenecks if they have adequate information on the state of the roads, traffic volume, and people's commutes. This, in turn, will lead to more sustainable and clean services and environments. Fig. 1 shows some of the most important AI applications in smart cities now in use and is followed by a brief description of each.

Healthcare:

This capacity to automatically analyse massive amounts of data, find hidden patterns, and derive clinically useful insights from that data is what drives ML applications in the healthcare industry. Medical professionals benefit from the automated extraction of insights because they save time and money while improving the quality of care for patients [28]. Recent advances in deep learning have allowed AI to be widely used in the healthcare industry and have shown it to be very beneficial. A method suggested byGoogle [29] surpassed human physicians (i.e., by roughly 16 percent accuracy) in the detection of breast cancer in mammograms. Both [30,31] and the skin cancer and lung cancer diagnoses made using AI techniques have been quite successful. Fig. 1 illustrates several intriguing AI applications in smart cities.AI can have a positive impact on transportation in a number of ways. When it comes to route optimization, for example, its predictive skills may assist estimate traffic flow and congestion. AI algorithms may also be used with multimedia processing approaches for road safety [33], driver distraction [34], and accident eventsdetection [35,37]. As a backbone for self-driving automobiles, artificial intelligence (AI) may be thought of as a system that continuously monitors its surrounding environment and makes predictions about future occurrences, including people, vehicles, and other roadside items [38].

Education: AI has a number of benefits for education, including automated grading andevaluation and prediction of students' graduation rates as well as personalised learning systems and intelligent tutoring systems. Predicting students' future career choices might potentially benefit from AIpredictive skills by using AI approaches to students' data, such as their interests and achievements in various areas.Another interesting smart city application where AI has shown its potential is crime detection, prediction, and tracking. The use of artificial intelligence (AI) in law enforcement is revolutionising the way agencies prevent, identify, and respond to crimes. Today's law enforcement organisations mainly depend on predictive analysis to monitor crimes and identify the most susceptible parts of a city, where more force and patrol teams may be placed. For example, PredPol uses artificial intelligence (AI) to anticipate "hot spot" crime areas.A clean and sustainable environment may also be monitored and maintained thanks to the aid of artificial intelligence (AI). [40] Environmental monitoring and enforcement have become more effective thanks to recent advances in deep learning and satellite technology. The use of artificial intelligence (AI) in environmental change analysis is widespread. Artificial Intelligence (AI) systems have also been shown to be very useful in disaster warning, water management, and waste sorting [41, 42, 43].In a smart building, a building's functions, such as lighting, heating, air conditioning, and security are controlled by an automated structure/system.The use of artificial intelligence (AI) in smart systems has been extensively documented in [44].[45] The tourism and entertainment sectors are also reaping the benefits of AI and social media. As an example, tourists utilise AI-based recommendation systems to help them choose their vacation locations, taking into account a variety of factors, including transportation and lodging facilities, culinary and historical places of interest as well as the cost of the trip. AI-based apps may also assist passengers in detecting fraud, reducing expenses, and locating entertainment and transportation options while they are on the road. Visual sentiment analysis tools powered by AI might be used to search or extract scenes from lengthy TV shows videos based on sentiment analysis [46], apart from recommendation systems, which is one of the key uses of AI in the field. In spite of its impressive results and success, artificial intelligence (AI) presents a number of issues, including privacy concerns and the potential for bias in public services. Data on travel patterns, for example, requires extensive collection and processing by the government, putting people's personal information at danger. It's much more harmful when AI makes judgments based on prejudice, whether it's deliberate or inadvertent, since it might put people's lives in risk in healthcare or law enforcement. According to a study, an AI-based programme used to forecast future offenders was shown to be prejudiced against blacks. Black individuals were assigned greater risk ratings (risk factors) by a smart system used to estimate health care requirements for nearly 70 million patients in the United States [47]. Algorithms don't learn prejudice on their own, but rather the data used to train them reflects the societal and institutional biases that have been adopted by society through the years [3]. As a result, AI algorithms mirror the values and priorities of people since they are created by humans (i.e., developers). For example, AI systems need correctly labelled and well represented training data in order to generate accurate predictions. If a class is overrepresented, this might lead to forecasts leaning toward the class. False positives and false negatives are also critical for AI predictions. These artificial intelligence (AI) limitations obstruct its ability to overcome societal and political prejudices and accomplish the genuine goals of smart cities. Artificial Intelligence algorithms in smart city applications are heavily impacted by the social and political decisions of the society and the government, according to Green [3]. In order to secure privacy and prevent prejudice in AI algorithms used in human-centric applications, we must first evaluate the necessity, aims, and possible influence of their judgments on society. Smart city applications also face a number of security vulnerabilities, such as adver-sarial assaults on AI models to influence the judgments by disrupting their ability to forecast the future. As an example, an attacker may disable an autonomous automobile on a busy highway and demand money to get it back up and running again. [48] A more dangerous issue would be if a train stopped on the platform shortly before the arrival of the following one. Another issue is that AI models' decisions are difficult for humans to decipher due to a lack of interpretability. The notion of explainability and ethics in AI has been established in order to mitigate the hazards associated with AI deployment in smart city applications. On this page, we'll discuss the ethical (data and algorithmic) issues that may arise while using artificial intelligence in smart cities.

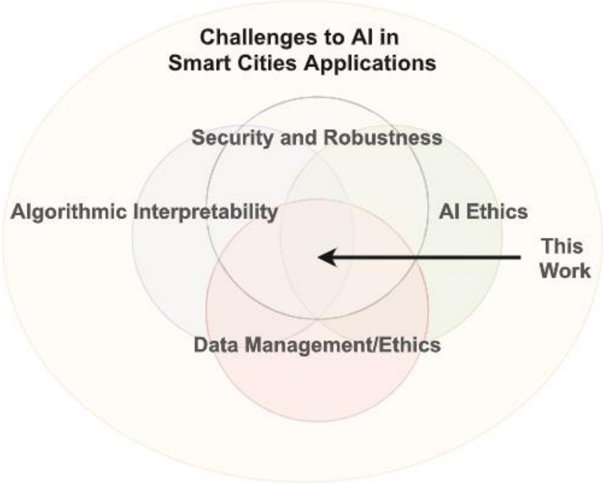

Challenges to AI in Smart Cities Applications
Security and Robustness
Algorithmic Interpretability
AI Ethics
This Work
Data Management/Ethics

Fig. 2:Visual depiction of the scope of the paper.

Scope of the survey

AI in smart city applications faces a number of hurdles, including security, robustness, interpretability, and ethical (data and algorithmic) issues. The paper's scope is shown in Fig. 2. Throughout the study, we examine how one of these challenges/problems might cause or assist in addressing another. We also keep tabs on the latest developments in these fields of study. It advises on the existing limits, hazards and future directions of research in various disciplines, and how it may fill the current gaps in the literature and lead to improved solutions, as well.

Related surveys

It has long been a hot field of study because of the research community's interest in ageing AI for smart city applications. Several intriguing publications have been published in the literature examining various elements of AI applications in smart cities [9]. There is also a considerable amount of material available on adversarial assaults, explainability, dataset availability and ethics in the context of human-centered AI. More attention has been paid to adversarial and explainable AI in the literature. Additionally, there are a number of intriguing polls on similar subjects that focus on various parts of the subject at hand. Nevertheless, to the best of our knowledge, no study has looked at all four issues at the same time, highlighting the interdependence of the many issues examined. An overview of adversarial attacks on deep learning models may be found in Zhang et al. [50]. As an example of onadversarial object recognition, see Serban et al. [51] for a complete review of the literature. For adversarial AI, Zhou et al. [52] present a review of game-theoretic techniques. A few recent polls on explainableAI have also been published. When it comes to ex-plainable AI, there is a survey in [53–56]. Explainable AI techniques are the topic of certain polls. There are a number of surveyweb and reinforcement learning-based methods to explainability, including [57] and [58]. A study of AI efforts on ethics, risk, and policy by Baum et al. [59] gives a more comprehensive picture of the current state of affairs. There is an overview of the literature on AI ethics in healthcare provided by Morleyet al. [60]. Instead of looking at the four difficulties separately, this study focuses on the connections between them and examines how a solution to one of the challenges may potentially benefit or harm the other challenges. 4

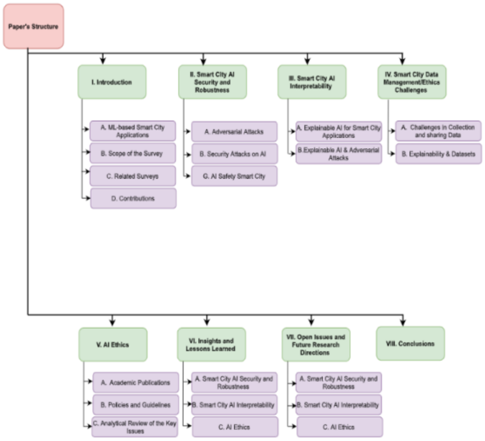

Paper's Structure
I. Introduction
A. ML-based Smart City Applications
B. Scope of the Survey
C. Related Surveys
D. Contributions
II. Smart City AI Security and Robustness
A. Adversarial Attacks
B. Security Attacks on AI
G. AI Safety Smart City
III. Smart City AI Interpretability
A. Explainable AI for Smart City Applications
B.Explainable AI & Adversarial Attacks
IV. Smart City Data Management/Ethics Challenges
A. Challenges in Collection and sharing Data
B. Explainability & Datasets
V. AI Ethics
A. Academic Publications
B. Policies and Guidelines
C. Analytical Review of the Key Issues
VI. Insights and Lessons Learned
A. Smart City AI Security and Robustness
B. Smart City AI Interpretability
C. AI Ethics
VII. Open Issues and Future Research Directions
A. Smart City AI Security and Robustness
B. Smart City AI Interpretability
C. AI Ethics
VIII. Conclusions

Fig. 3. Structure of the survey.

Contributions

A comprehensive review of the literature on the security, safety, robustness and interpretability of AI in smart city applications is presented in this study. These ideas and difficulties are intertwined and interdependent in the study, which examines the relationship between them. The following are the paper's primary contributions:A thorough examination of AI's role in the development of modern cities, as well as its possible problems, such as ethical, interpretation, security and fairness issues that might hamper its progress in various smart city applications, is provided in this paper..Security, safety, robustness, interpretability, and ethical concerns in implementing AI in human-centric applications are all examined in the article.Moreover, the report gives important insights into the link between these difficulties and illustrates how they may impact one other.

• We also point out the hazards, limits, and unsolved research questions in these areas.

Here are the sections of the paper. AI security and robustness concerns in smart city applications are discussed in Section 2. Explainable AI, as discussed in Section 3, is critical to gaining a deeper understanding of AI choices and may be connected to adversarial assaults. Section 4 discusses the difficulties in collecting and distributing data. Using AI in human-centric smart city applications has ethical implications, which are discussed in Section 5. Section 6 presents the most important findings and takeaways from the research. In Section 7, we highlight the outstanding topics and research avenues that are yet to be explored. There are some final observations in Section 8.Security and robustness of AI in smart citiesIn smart cities, machine learning has the potential to increase the efficiency and productivity of many municipal processes. AI in smart cities has produced excellent results, but security is still a major challenge that needs additional exploration and experimentation. It is possible for several types of assaults against AI models to be used against them; they include: adversarial examples; model extraction; backdooring; trojans; membership inference; model inversion Cyberattacks on AI models place new demands on software security systems and methods that must be adapted to deal with a new set of problems [62]. One of the most dangerous things about AI is the fact that even a little adjustment to inputs or data absorbed by the algorithms may have a significant impact on the AI models' decisions. It is obvious from the following selection of AI security challenges that a thorough examination of AI safety and security while converting cities to be smart is urgently needed. As of 52016, Tesla's Auto-driver feature also caused confusion among2016: white truck side with sky, heading to

There was a fatal accident.1After only a few hours of operation in 2016, Microsoft's chatbot was shut down and shuttered. The model was assaulted and made to tweet degrading messages for the benefit of the public. 2 Using chatbots for government services isn't only for companies. Citibot, a communication tool for residents and their governments, was recently established by the US city of North Charleston in South Carolina. Requests for information or repairs may be made by citizens. 3 In order to assess and handle citizen demands, these smart systems rely on people' data.2016 Google AV was in autonomous mode when it crashed due to a problem with speed estimate.

Attacks against facial recognition systems in 2016 employing eyeglassesframes [63].

A inexpensive 3D-printed mask deceived Apple's face-recognition system in 2017.

In 2018, a pedestrian was murdered by an Uber self-driving vehicle. There is an issue with the AV's timing.

- In 2018, a self-driving car's DNN classifier might be fooled into misclassifying speed limit signs [64].

DeepSpeech, a deep learning-based speech recognition system, has been effectively exploited by targeted audio adversarial instances in 2018. It is possible to transmit secret directives by the use of sound waves [65].Tiny adjustments to lane markers at Tencent's Keen Security Lab in 2019 successfully targeted Tesla's AI-powered autopilot system. Model S swerves to the wrong lane, making Tesla's lane-recognition models unsafe and inaccurate in particular situations.Neural networks have been used to the diagnosis of benign moleas in 2019. The medical picture has been tainted with cancerous cells due to the addition of microscopic noise [67].This year, Deepfakes. Deepfakedetection uses data from a collection that Facebook has gathered. 7

As of 2019, the algorithm that guides treatment for millions of individuals in New Jersey, USA, is prejudiced towards dark-skinned patients. For individuals with the same medical issues, dark-skinned people are given lower ratings. 8

The next year, 2020, Fuzhou Zhongfang Marl-boro Mall in China has a shopping guide robot that walks to the escalator by itself. Passengers were knocked down when the escalator toppled. The robot has been taken off the job indefinitely. 9

Due to a problem with the self-driving software on roads, Starsky Robotics will be shut down in 2020. Reported by the Starsky team, supervised ML isn't all very great.

Deepfake content, based on AI, is expected to be more prevalent by 2021 according to an FBI alert.

For making rude remarks and disclosing user information, the AI chatbot was banned in 2021 and the company is now being sued in South Korea.

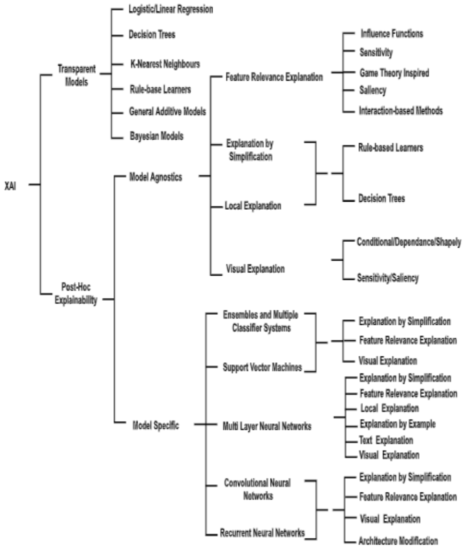
It's clear from this list that there are a number of challenges that need to be addressed in addition to developing AI models that perform well. Attacks on machine-learning models in smart city applications are discussed in the following sections, which highlights the need for secure and resilient AI solutions at technical and policy levels.

Adversarial attacks

When it comes to creating ad-versarial instances, this difficulty has been acknowledged and addressed either for creating false data that might belong to many domains, such as text, photos, audio, and network signals, or analysing and finding solutions to this security concern [71]. A model M may classify some data as class 1 if it comes from a benign source, but it will classify other data as class 2 if it comes from a poisoned function F, which would produce data such as X ′, which is not detectable by humans. The term "untargeted adver-sarial assaults" refers to this kind of attack. As a result of the targeted assaults, the model M gets biassed towards Y in prediction since function F is attempting to identify alternative benign input. Adversarial assaults may also be classified according to how much information the attackers have about the target model (victim). White-box, gray-box and black-box threat models are all examples of a threat model. adversary have complete knowledge of the targeted model architecture in white-box assaults. As a result, the system is able to be fooled by poisoned data. Threat models with a grey box may provide attackers access to the model's general structure, whereas those with a black box have no such information [72]. Adversarial assaults are classified in great detail in [5]. An image classification classifier may be tricked by introducing a minor perturbation, increased in the picture for visual illustration, to a valid sample to disrupt the prediction capabilities of the model. In the next paragraph, we present an overview of the literature on adversarial assaults on AI models in smart city applications without delving into more depth. 12 https://securityintelligence.com/articles/how-protect-against-deepfake-attacks-extortion/.13 https://www.straitstimes.com/.6

Smart city AI interpretability

As is the case with most AI frameworks, a collection of characteristics is fed to an AI algorithm that then uses the data to find patterns and makes predictions. There is no explanation or rationale for the predictions in these "black-box" frameworks, which are also known as "black sandboxes." An explainable AI framework, on the other hand, provides information on the reasoning behind the prediction/decision in addition to its predictions/decisions. Additional functions/interface are used to comprehend the underlying reasons of a choice [56] in order to achieve this goal. Interpretability and explainability are often used interchangeably in the literature. It's only that there are a few differences in terminology. While interpretability demonstrates the amount to which a system can be seen, explainability illustrates the extent to which AI algorithms can be explained to a human. AI models are used to make critical decisions about people's lives, such as whether they get a job or not (AI-based recruitment), or whether an individual is guilty/involved in a crime or not (i.e., pre-trial decision-making). These justifications of the causes of an AI model's predictions may be gained in two ways, according to Guidotti et al. [142], which they referred to as "Black-boxExplanation' or directly creating and implementing transparent AIalgorithms. Explanable artificial intelligence has the following advantages:


XAI
Transparent Models
Logistic/Linear Regression
Decision Trees
K-Nearest Neighbours
Rule-base Learners
General Additive Models
Bayesian Models
Post-Hoc Explainability
Model Agnostics
Feature Relevance Explanation
Influence Functions
Sensitivity
Game Theory Inspired
Saliency
Interaction-based Methods
Explanation by Simplification
Local Explanation
Rule-based Learners
Decision Trees
Visual Explanation
Conditional/Dependance/Shapely
Sensitivity/Saliency
Model Specific
Ensembles and Multiple Classifier Systems
Support Vector Machines
Explanation by Simplification
Feature Relevance Explanation
Visual Explanation
Multi Layer Neural Networks
Explanation by Simplification
Feature Relevance Explanation
Local Explanation
Explanation by Example
Text Explanation
Visual Explanation
Convolutional Neural Networks
Recurrent Neural Networks
Explanation by Simplification
Feature Relevance Explanation
Visual Explanation
Architecture Modification

A taxonomy of achieving transparent and explainable AI decisions by opening the so called black-box AI models [56].

Various criteria may be used to identify explainable AI approaches [56,143,144]. Taxonomy of explainable AI may be seen in Fig. 9. Transparent models and post-hocexplainability are the two basic kinds of ex-plainable AI. For example, a model's complexity may be limited for the sake of explaining it, while the model's behaviour can be analysed after it has been trained. A trade-off exists between performance (e.g., accuracy) and explanation, as should be mentioned. Transparency models like fuzzy rule based predictors have been shown to be less accurate than so-called black box approaches like CNNs in the literature [145]. However, in critical applications, such as healthcare, smart grids, and predictive policing, explanation and interpretability are desired features. In this regard, the development of post hocexplainable procedures is a special emphasis. Explainable AI models are critical in smart city applications, and we examine how they deal with adversarial assaults and the ethical issues that arise from using explainable AI.

Explainable AI for smart city applications

There are several benefits to AI that may be attributed to explainability [55,146,147]. Because of the technology's direct influence on society and its residents, smart city applications can't ignore its significance. Some of the most important smart city applications, such as health care and transportation as well as banking and other financial services, necessitate the use of explainable artificial intelligence (AI). This includes decisions about humans, such as who should receive a particular service; which medicine should be used; and who should be employed. For improved predictions and judgments in smart city applications, it is necessary to evaluate data (i.e., features) in order to remove any inaccuracies [148]. Health care is one of the most important uses of smart cities, and it necessitates AI models that can be explained rather than conventional black boxes. It is undeniable that artificial intelligence (AI) has been a boon to the healthcare industry, yet typical black-box AI just makes judgments without interpretation. Explainable AI in healthcare is needed for a variety of reasons, including the wide-reaching repercussions and high costs involved with a miscalculation [149]. Furthermore, boosting physicians' confidence in AI-based diagnosis requires an in-depth knowledge of the reasons of AIpredictions and choices. It is easier for doctors to make judgments based on an AI diagnosis if the AI model's conclusion can be understood and interpreted by humans. Additionally, the domain experts' expertise may be used to enhance AI models that are easier to understand. Also in healthcare, predictive performance is not adequate to get clinical insights for making judgments [150]. There are seven pillars of explainable AI in healthcare in [142,151]: transparency, domain sense, concision, generalizability, trust/performance and integrity. These pillars indicate how AI in healthcare is related to these characteristics. Another major use of smart cities is transportation and autonomous vehicles, where mistakes made by an AI model have severe effects and costs connected with them. Errors like as failing to distinguish correctly between red and green traffic signals, or failing to recognise pedestrians, may result in serious harm to people's lives and property. Due to a prediction/classification mistake, a self-driving Uber in Arizona struck and killed a woman. [55,169] This has occurred before. A woman was struck and killed by a self-driving Uber. We11

5.3.1. Singularity/superintelligence

Artificial intelligence (AI) advancements are seen as a danger to humankind's very existence, rather than a threat to individual health or wellness. Concern about the "singularity theory," as it is known, is common. If AI systems succeed in creating machines or robots with an intellect comparable to that of humans, then these machines will be able to act independently and develop their own "superintelligent" machines, which will ultimately be more intelligent than humans themselves. The point of "singularity", as in physics, will be a logical result of such a shift-making sequence of advancements. It will be impossible for humans to influence anything at this point, even their own destiny, therefore the superintelligent machine will be the last human invention. As a result, our current conceptions of human affairs and fundamental values (perhaps even what it is to be human) will be shattered. 18 There are some people who think that humans will be supplanted by superintelligent robots in a post-singularity future, and humanity will therefore become obsolete. There are many who believe that humans will evolve into superhuman creatures rather than being annihilated. Humans will be able to achieve ever-increasing levels of intellect, capacity, and longevity via the process of mutual hybridization between humans and robots [250,312]. It's also possible that some voices think the singularity hypothesis is bogus; others argue that it overestimates the dangers of artificial intelligence. Because of this, some have questioned whether or not this idea is worthy of being taken seriously in moral debate, or whether it belongs in the realm of science fiction instead [250,311,312]. Singularity opponents have even accused its proponents of lacking AI job experience [206,295]. Many of the above-mentioned rules and recommendations may have been silent on the singularity theory because of their scepticism about its seriousness [302]. However, despite the word "singularity" appearing in the title of a 2017 paper from the US Center for a New American Security (CNAS), no substantial examination of the singularity theory was provided[313]. For example, when "Preparing for the Future of Artificial Intelligence" cited the "singularity hypothesis," it was said that this hypothesis should have minimal influence on present policy and should not be the primary driver of AI publicpolicy." Ethically aligned design, in its earlier iteration, took a similar stance, warning against adopting "dystopian assumptions about autonomous machinesthreatening human autonomy" [306] in reference to the singularity theory. 5.3.2. The branch of human interest (AI ethics) Concerns around the use of personal information (e.g., privacy, transparency, explainability, ad-versarial attacks). In general, the quality of the training data has a significant impact on the effectiveness of AI systems. A significant portion of the AI ethical difficulties and dilemmas focus on the subject of how massive data should be handled in an ethically sound manner. The AI systems are really executing a modernised version of the traditional state surveillance by secret services while aiming to gather and analyse as much data as possible. Face recognition and device fingerprinting, as well as "smart" phones and TVs, "smart governance," and the "Internet of Things" are just a few of the data-gathering methods that may be employed in smart cities. According to some observers, these instruments will be able to learn more about us than we know about ourselves. Consequently, one may utilise the information acquired to influence one's actions. Other than being used to violate people's privacy and information secrecy, this vast data collection machine has the potential to profit from our obtained data without our knowledge or agreement. "Surveillance economy" and "surveillance capitalism" are other names for this phenomenon. [250,314,315]. Section 4 has a more in-depth treatment of data-related ethical considerations, such as privacy, bias, ownership, data openness, interpretation, and informed permission. Another crucial feature of big data management is explainability, which is directly tied to moral concepts such as justice, bias, responsibility, and trust. The idea of trans-parency, which would simply involve producing a readily understandable overview of system operation, interacts with the minimal degree of explainability. The AI systems must keep detailed records of when, how, who, and why they were built. These records must also be explicable and understandable to the general public. There are a number of ways in which AI systems may be set up to collect and store this data. Technically speaking, explainability involves additional moral constraints. Meaning that humans should be able to comprehend and explain the reasoning behind an AI model's decisions in order to be aware of the model's potential biases and their possible causes [317]. Public and academic disputes concerning probable moral transgressions relating to discrimination, manipulation, prejudice, unfairness, etc. continue to be triggered by the lack of explainability and transparency, which will be seen as opacity. [318] A Goldman Sachs AI system was said to be discriminatory against women. There were claims that the Google Health research, which claimed an AI system could beat radiologists in the prediction of cancer, was unreliable [319–321]. Artificial Intelligence (AI) researchers are working to solve these concerns by creating approaches that allow for the so-called "explainable AI," as well as "discrimination-aware datamining." Although governments continue to encourage the AI sector for more explainable applications, the opposite is also true. According to the EU's GDPR, individuals have the right to know why their personal information is being collected and used. [317].

References

[1] Secure, sustainable smart cities and the IoT, 2020, Accessed: 2020-06-24,https://tinyurl.com/y6qw479s.

[2] A. Gharaibeh, M.A. Salahuddin, S.J. Hussini, A. Khreishah, I. Khalil, M.Guizani, A. Al-Fuqaha, Smart cities: A survey on data management,security, and enabling technologies, IEEE Commun. Surv. Tutor. 19 (4)(2017) 2456–2501.

[3] B. Green, The Smart Enough City: Putting Technology in Its Place toReclaim Our Urban Future, MIT Press, 2019.

[4] ARUP: If you know the right questions and understand the risks, datacan help build better cities, 2020, Accessed: 2020-07-07, https://tinyurl.com/y4b8bq6e.

[5] A. Qayyum, J. Qadir, M. Bilal, A. Al-Fuqaha, Secure and robust machinelearning for healthcare: A survey, 2020, arXiv preprint arXiv:2001.08103.

[6] M. Veres, M. Moussa, Deep learning for intelligent transportation systems:A survey of emerging trends, IEEE Trans. Intell. Transp. Syst. (2019).[7] J. Xie, H. Tang, T. Huang, F.R. Yu, R. Xie, J. Liu, Y. Liu, A surveyof blockchain technology applied to smart cities: Research issues andchallenges, IEEE Commun. Surv. Tutor. 21 (3) (2019) 2794–2830.

[8] K. Ahmad, J. Qadir, A. Al-Fuqaha, W. Iqbal, A. El-Hassan, D. Benhaddou,M. Ayyash, Artificial intelligence in education: A panoramic review, 2020.

[9] Z. Ullah, F. Al-Turjman, L. Mostarda, R. Gagliardi, Applications of artificialintelligence and machine learning in smart cities, Comput. Commun.(2020).

[10] S. Latif, A. Qayyum, M. Usama, J. Qadir, A. Zwitter, M. Shahzad, Caveatemptor: the risks of using big data for human development, IEEE Technol.Soc. Mag. 38 (3) (2019) 82–90.[11] H. Ekbia, M. Mattioli, I. Kouper, G. Arave, A. Ghazinejad, T. Bowman, V.R.Suri, A. Tsou, S. Weingart, C.R. Sugimoto, Big data, bigger dilemmas: Acritical review, J. Assoc. Inform. Sci. Technol. 66 (8) (2015) 1523–1545.[12] K. Crawford, R. Calo, There is a blind spot in AI research, Nature 538(7625) (2016) 311–313.[13] Machine bias: There's software used across the country to predict futurecriminals. And it's biased against blacks., 2020, Accessed: 2020-08-26,https://tinyurl.com/j847koh.[14] K. Crawford, Artificial intelligence's white guy problem, N.Y. Times 25(06) (2016).

[15] A. Qayyum, M. Usama, J. Qadir, A. Al-Fuqaha, Securing connected &autonomous vehicles: Challenges posed by adversarial machine learningand the way forward, IEEE Commun. Surv. Tutor. 22 (2) (2020) 998–1026.

[16] S.C.-H. Yang, P. Shafto, Explainable artificial intelligence via Bayesianteaching, in: NIPS 2017 Workshop on Teaching Machines, Robots, andHumans, 2017, pp. 127–137.

[17] S.M. Lundberg, B. Nair, M.S. Vavilala, M. Horibe, M.J. Eisses, T. Adams, D.E.Liston, D.K.-W. Low, S.-F. Newman, J. Kim, et al., Explainable machine-learning predictions for the prevention of hypoxaemia during surgery,Nat. Biomed. Eng. 2 (10) (2018) 749–760.

[18] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra,Grad-cam: Visual

explanations from deep networks via gradient-basedlocalization, in: Proceedings of the IEEE International Conference onComputer Vision, pp. 618–626.

.

[19] Amazon scraps secret AI recruiting tool that showed bias against women,2020, Accessed: 2020-08-26, https://tinyurl.com/y8eelatr.

[20] The two-year fight to stop amazon from selling face recognition to thepolice, 2020, Accessed: 2020-08-05, https://tinyurl.com/y8q7cvue