GMR
The Original

EXPLAINABLE ARTIFICIAL INTELLIGENCE AND EDGE COMPUTING FRAMEWORKS FOR SECURE DECISION-MAKING IN SMART DIGITAL ECOSYSTEMS

Dr. Teena[1*], Dr Surekha Manohar Jadhav[2], Sunil R. Gagare[3], S. N. Timmanagoudar[4], Dr Vaibhav Panwar[5], Ranjan Banerjee[6]

[1*]Associate Professor, Maharishi University of Information Technology, Email ID: tina.summer27@gmail.com ORCID: 0000-0002-3378-3051
[2]Assistant Professor, Bachelor of Computer Applications, Women's College of Home Science and BCA, Loni, Email ID: surekha.kale@pravara.in ORCID: 0009-0005-9499-4302
[3]Assistant Professor, Department of Electronics & Telecommunication Engineering, Amrutvahini College of Engineering, Sangamaner, Savitribai Phule Pune University, Pune, Email ID: sunilgagare@yahoo.co.in, ORCID: 0000-0003-3239-3569
[4]Assistant Professor, Department of Mechanical Engineering, K.L.E. Institute of Technology, Gokul Hubballi-580027, Email ID: shrinivasnt@kleit.ac.in ORCID: 0009-0008-1495-0687
[5]Associate Professor, Department of Center for Distance and Online Education, Specialization in Data warehousing and Data mining, Manipal University Jaipur, Email ID: Vaibhav.panwar@jaipur.manipal.edu
[6]Assistant Professor, Computer Science & Engineering, Brainware University, Barasat, Kolkata – 125, West Bengal,700125, India, Email ID: rbkpccst@gmail.com ORCID:0009-0003-1950-7530

ABSTRACT
Artificial Intelligence is increasingly a key enabler of decision-making, with fast, adaptive, data-driven approaches to smart digital ecosystems in healthcare, smart cities, Industry 5.0, energy systems, IoT networks, and cyber-physical infrastructures. The increased and potential deployment of opaque AI models in distributed settings, however, comes with concerns about trust, security, accountability, privacy, and operational resilience. This review examines the convergence of explainable artificial intelligence and edge computing as a foundation for secure decision-making in smart digital ecosystems. It synthesizes the conceptual foundations of smart ecosystems, edge intelligence, trustworthy AI, and cybersecurity, followed by an analysis of major XAI techniques, including intrinsic models, SHAP, LIME, counterfactual explanations, attention-based methods, and knowledge graph-supported approaches. The review further discusses edge computing frameworks based on IoT–Edge–Cloud integration, federated learning, privacy-preserving computation, blockchain-enabled trust, and security-aware decision mechanisms. Real-world implementations are examined across critical application domains to show how explainable edge intelligence can improve transparency, responsiveness, and human confidence. The analysis identifies persistent challenges related to resource-efficient explainability, adversarial robustness, privacy preservation, model validation, governance, and context-aware deployment. In conclusion, this review underscores the importance of developing common standards and guidelines that incorporate explainability, security, edge intelligence, and human-centric governance to support clear and secure decision-making in the smart digital ecosystems of the future.

KEYWORDS: Explainable Artificial Intelligence, Edge Computing, Secure Decision-Making, Smart Digital Ecosystems, Trustworthy AI

1. INTRODUCTION

Smart digital ecosystems have proven to be environments of smart technologies, digital platforms, cyber-physical systems, and data services, where the technologies, platforms, systems, and services are all interconnected so that a seamless interaction between the users, devices, and organizations is achieved. These systems continue to be digitized, changing the traditional way of doing business: adaptive, collaborative and autonomous capabilities are developed in various fields, from healthcare to transportation, manufacturing to urban infrastructure [1]. The transformation is further reflected in the emergence of smart cities in which the combination of connected intelligence, platform-based services and network effects enables large-scale decision-making processes and the optimization of services [2]. The complexity and scale of a digital ecosystem are growing, and intelligent decision-making is a vital need for success in terms of efficiency, responsiveness, and sustainability.

With large, complex, and varied datasets, artificial intelligence (AI) has emerged as a critical tool for decision support in today's digital ecosystems by providing the ability to analyze, model, and understand the data and offer actionable insights for complex operations. Data-centric reasoning and predictive analytics have been found to have significant potential in enhancing strategic planning, resource allocation, risk evaluation, and organizational performance, thanks to AI-driven decision support systems [3]. In the era of AI technologies, decision-making paradigms have increasingly become complex, such as the use of expert systems, recommender systems, and explainable AI (XAI) with the aim of improving the accuracy and transparency of automated decisions [4]. However, the challenges that arise due to such progressions

involve the issues of interpretability, accountability, and trust in the use of the AI model, especially when it comes to decision-making for essential societal and industrial tasks.

The enormous quantity of data that is being generated through the widespread usage of the Internet of Things (IoT) and real-time applications has made it so that there is more data than ever before, which limits. To solve this problem, edge computing has evolved as a viable solution, which moves the computational intelligence closer to the data sources, thus decreasing latency, bandwidth consumption and response time. The concept has been extended with Edge Intelligence, by taking advantage of AI at the edge to enable localized analytics and decision making at the edge [5]. Cloud Analytics frameworks based on Artificial Intelligence capabilities are still providing the required computational power to enterprise decision systems without compromising the responsiveness at the edge [6] and are creating a hybrid intelligence ecosystem that combines cloud resources with edge intelligence. This type of integration is gaining significance to enable time-critical applications to make quick decisions and make effective use of resources.

There is still a long way to go in terms of trust and security in smart digital ecosystems, as the trust of distributed generation, processing and communication of data is still a major challenge. There are multiple stakeholders and multiple components (both devices and software) that create complicated relationships of trust that directly impact the reliability of the system and the user trust [7]. Meanwhile, there are emerging cyber threats to digital infrastructure, IoT networks and AI-powered systems, leading to a heightened awareness of the need for data privacy, system integrity, and operational resilience [8]. All the more significant are these challenges when decisions are made by autonomous AI systems without clear explanations of how they make the decisions.

Previous studies have investigated the application of the AI systems as decision makers, but Explainable AI, Edge Intelligence and Cybersecurity are still relatively separate fields and limited research has been conducted on integrating them in the context of a smart digital ecosystem. The current literature, nevertheless, does not offer an overview that provides a comprehensive outlook on the opportunities and potentials of using the tandem of explainable AI and edge computing architectures from the standpoint of security, transparency, and trustworthy decisions in the context of distributed digital network environments. This review is undertaken to fill this gap through providing a critical summary of the current state-of-the-art in the area of explainable AI and edge computing for secure decision-making of smart digital ecosystems. The goal of this research is to conduct a survey of existing architectures as well as to gain insight into explainability approaches, security methods, applications, and new research directions in the area of trusted and resilient ecosystems.

2. REVIEW METHODOLOGY

The thorough review was conducted to identify and summarize literature sources about explainable AI and secure decision-making in intelligent digital ecosystems within the framework of edge computing. Articles chosen were assessed according to their relevance to key topics of the paper including smart ecosystems, edge computing, XAI approaches, federated learning, cybersecurity measures, trustworthy AI and their practical applications. The literature was grouped by topics related to conceptual and technological, security, application and future studies aspects. Publications where either theoretical insight, methodology, architecture or practical application could be regarded as top priorities. The similarities, problems and unresolved issues of reviewed articles were compared critically. Such an approach helped to provide a general yet in-depth perspective on the role of explainable edge intelligence in protecting digital trust and security.

3. Foundations of Explainable AI, Edge Computing, and Secure Smart Digital Ecosystems

Technological innovations in intelligent systems are progressing fast and giving rise to the creation of smart digital ecosystems, which involve the coming together of inter-connected devices, computing infrastructure, data platforms, and autonomous services in one single operating system. These systems provide smooth interaction between the physical world and the digital world, thereby allowing efficient decision-making, effective service delivery in various domains, and increased efficiency. The major elements of these ecosystems are smart systems, which help form a network of inter-connected systems capable of sensing, processing, and responding to their dynamic environments. They are capable of transformation, but these ecosystems encounter challenges in terms of interoperability, scalability, governance, and sustainability. Digital twin technologies, which simulate physical systems and processes, have been a key addition to the smart digital ecosystems. Digital twins provide a means of bringing together, in a single analytical environment, a variety of data streams, including real-time data, which can be monitored, simulated, predicted, optimized and then optimized again. Digital representations of earth systems are excellent examples of how new architectures, centered around ecosystems, can enable complex data integration and intelligent decision support on distributed architectures [10]. As digital twins become more widely used, it becomes even more crucial to have computational frameworks that can handle the volume of data generated and the need for real-time analytics.

Edge computing has become a key architectural trend to meet the increasing needs for computing in today's smart ecosystems. Many time-sensitive applications are unable to meet the latency, bandwidth and responsiveness requirements of traditional cloud-centric solutions. Thus, the IoT–Edge–Cloud continuum has become a distributed computing model that is gaining much attention for its ability to support seamless orchestration of resources across a multitude of computational layers. Within this continuum, advanced frameworks that facilitate integration, software orchestration and service management create the basis for efficient deployment of intelligent applications and enable flexibility and scalability across heterogeneous environments [11]. With the increase in the number and speed of devices that are connected, it is becoming increasingly important for this transition to take place. It becomes vital for the IoT edge devices that are resource-constrained to process information smartly and effectively in terms of their processing, memory and energy capacity. Therefore, context-aware intelligence becomes a necessity in the edge environment as they can react to

their context and user requirements. This distributed intelligence enables fast decision-making in real time, while minimizing the reliance on centralized systems, enhancing system responsiveness and operational resilience [12]. Incorporating AI features in the edge environment has therefore led to the rise of edge intelligence, where the intelligent systems perform the computational decision-making at the edge, closer to the data source and end-user.

Since AI systems are increasingly taking part in various decision-making processes, transparency and interpretability are a major research priority. Explainable AI aims to make AI-driven decision-making understandable for the human stakeholders by giving them meaning-inspiring explanations of the behavior of the model, its predictions, and the reasoning processes. Research on XAI today includes many different methods designed to make an AI more transparent without significantly reducing its predictive accuracy [13]. The concepts behind explainability have been decades in the making, with the objective to gain insight into machine reasoning starting from the rule-based systems and moving forward to the modern methods of explainable and post-hoc machine learning explanation techniques for complex machine learning models [14]. In cases where trust, accountability, and compliance are required, explainability becomes ever more important. With the rising amount of personal information being shared in the smart digital ecosystem, there is now a huge requirement for maintaining user privacy and data governance & compliance. In privacy-first, the emphasis is on how security protocols have been used to ensure that user information is protected while also complying with all regulations and smooth functioning [15]. Today, Trustworthy AI is a holistic concept that not only addresses data protection but also transparency, fairness, robustness, accountability and reliability of the AI lifecycle. This is an important point to bring out from a computational point of view of trustworthy AI [16].

Explainability of AI, edge computing, and cybersecurity are some of the key issues in next-generation smart digital ecosystem. Security interactions between the edges of devices and cloud-to-cloud, as well as cloud-to-device interactions, become critical in ensuring continuity of operations and infrastructure resilience. The deployment of 5G technology and other advanced communications technologies, including edge computing and cloud security, become one of the key methods of enhancing the resilience of critical digital infrastructure against constantly evolving cyber threats [17]. All these key technologies can be combined together in order to ensure safe, secure and transparent decision-making in smart digital ecosystems. To get a better understanding of the conceptual basis of this review, Table 1 presents a summary of the key technological layers that link explicable AI, edge computing, security and smart digital ecosystems.

Table 1. Foundational Components of Secure Explainable Edge Intelligence

Component	Core Role	Key Technologies	Main Benefit	Key Challenge
Smart Digital Ecosystems	Connect physical, digital, and organizational entities	IoT, CPS, digital twins, smart platforms	Integrated and adaptive decision-making	Interoperability and governance
Edge Computing	Processes data close to the source	Edge nodes, fog layers, gateways	Low latency and reduced bandwidth use	Limited computation and energy
Explainable AI	Makes AI decisions understandable	Interpretable models, SHAP, LIME, counterfactuals	Transparency and accountability	Explanation fidelity and usability
Secure AI	Protects data, models, and decisions	Encryption, authentication, anomaly detection	Resilience against cyber threats	Adversarial and privacy risks
Trustworthy Intelligence	Aligns AI with human and regulatory expectations	Governance, auditing, fairness, robustness	Reliable and responsible automation	Validation across dynamic contexts

4. Explainable AI Techniques for Edge-Based Decision-Making

The transparency, accountability and operational trustworthiness of the edge-based decision-making process are crucial to XAI techniques. Edge AI systems are frequently used to generate decisions near the data sources, which could be IoT sensors, mobile devices, embedded controllers or industrial gateways. This localized intelligence enhances the responsiveness, and also implies the need for explanation, which can be achieved within the bounds of latency, energy, memory and capacity. Explainability and interpretability in edge AI cannot only be addressed at the model level, but must be also thought of as system-level requirement, depending on deployment context and resources, as well as user requirements and decision criticality [18]. One key difference between the XAI approaches is whether the explanations are generated during the model's execution (intrinsic) or after the fact (post-hoc). Intrinsic approaches strive for simpler models that have more direct explanations of their logic, e.g., decision trees or rule based models, linear models, symbolic reasoning systems and transparent feature based classifiers. These techniques may provide quick explanations with limited resources, and are suitable for giving explanations to others, especially in edge scenarios. Post-hoc methods make explanations after training the model, on the contrary and are often used for complex black-box models. In the context of smart manufacturing, both intrinsic and post-hoc XAI techniques have been applied to facilitate fingerprint identification

in operational scenarios and to predict responses, highlighting the role that explainability can play in technical diagnosis and human decision making in operational scenarios [19]. The principles of XAI highlight that accurate predictions are not sufficient for AI systems to be trusted in their decision-making process. Interpretability is aimed at understanding the model behavior, and explainability is aimed at conveying the model behavior in a meaningful form to the stakeholder. This distinction is worthwhile for making decisions at the edge as explanations could be used by engineers, operators, clinicians, security analysts or automated supervisory systems. A transparent decision-making framework should thus match the type of explanation with the knowledge and experience level of the user, risk and the urgency of the decision [20].

The model-agnostic explanation methods are important in the context of many edge applications where heterogeneous AI models are used. SHAP, LIME, and counterfactual analysis are methods that can be used by decision-makers to explore the significance of particular features and the local behavior of the model, as well as what could alter the results of a model. These techniques can help provide explanations to actionable insights in maintenance decision-making, such as highlighting the influential variables for equipment risk, failure prediction or prioritization for maintenance intervention [21]. These techniques are useful in a smart digital ecosystem to link algorithmic output to operational action. The explainable challenges with deep learning models also arise in the edge environment because of the complex internal representation of the models. Commonly used popular CNN models are convolutional neural networks, recurrent neural networks, long short-term memory networks and gated recurrent units for image, sequence, temporal and sensor-based data analysis. Their representational power, however, can simultaneously restrict direct interpretability and thus require explanation mechanisms when they are put into use in safety-critical or real-time applications on the edge side [22]. The challenge here is to maintain prediction accuracy while providing explanations that are computation-feasible at the edge. Explainability becomes even more complex when implementing federated and distributed learning, when the models are trained on several devices without having to be stored on a central server, where the raw data would also be stored. There is an avenue of research using human-in-the-loop (HIL) with domain-knowledge graphs that is well on the road to explainable federated deep learning, where domain knowledge is incorporated into distributed model learning and interpretation [23]. This is particularly crucial for the smart digital ecosystems with respect to privacy, decentralization, and domain-specific reasoning working collaboratively. One of the challenges that has arisen in all XAI techniques is the performance–interpretability tradeoff. A highly interpretable model might be able to provide good explanations, but with poorer predictive ability, whereas a more complex model might be able to be more accurate, but less interpretable. Comparative experiments of machine learning models indicate that careful selection of the model is important, as it must be reliable, interpretable, computationally inexpensive and carry the minimum risk of application [24]. Thus, incorporating explainability into edge-based decision-making frameworks needs to be an integrated part of the design, deployment and decision governance process, rather than an afterthought. Table 2 shows a comparison of the major explainable AI techniques for decision-making at the edge, and associate each technique with its role in explaining the decision, its suitability and its limitations in practice.

Table 2. Comparison of XAI Techniques for Edge-Based Decision-Making

XAI Technique	Explanation Type	Edge Suitability	Decision-Making Value	Limitation
Decision Trees and Rule-Based Models	Intrinsic	High	Clear logic for fast local decisions	Lower accuracy for complex data
LIME	Post-hoc, local	Moderate	Explains individual predictions	Sensitive to perturbation settings
SHAP	Post-hoc, local/global	Moderate	Quantifies feature contribution	Computationally expensive
Counterfactual Explanations	Post-hoc, actionable	Moderate	Shows how decisions could change	Requires realistic alternatives
Attention-Based Explanations	Model-specific	Moderate	Highlights influential inputs	May not always reflect true reasoning
Knowledge Graph-Based XAI	Hybrid/contextual	Emerging	Adds domain knowledge to explanations	Requires expert-curated knowledge

5. Edge Computing Frameworks for Secure and Explainable Intelligence

Edge computing frameworks are the architecture that enables the deployment of secure and explainable intelligence in distributed smart digital ecosystems that are distributed. The IoT ecosystem is characterized by smart decision-making that is spread across a variety of layers, from end devices to edge gateways, fog nodes to cloud platforms and everywhere in between. Explainable AI for IoT emphasizes the importance of embedding transparency mechanisms within such distributed systems, enabling humans and other systems to interpret, validate and trust the decisions made by AI systems [25]. Edge based intelligence is often real time and transparency is a critical element as there will be minimal accountability and confidence in key applications.

Distributed machine learning has proven to be a key component to the realization of intelligent decentralized infrastructures. Distributed learning entails the distribution of computing and/or model training process rather than data

collection in a centralized manner. Based on the concept of distributed learning, Federated learning facilitates the development of the model in a decentralized manner from one device to the other without transferring raw data between the devices, thereby promoting the privacy and efficiency of communication [26]. It is very relevant in the context of intelligent digital ecosystems since it provides data-driven applications without interfering with the autonomy of edge devices. Distributed intelligence adds complexity in the wireless and communication-constrained environment. The problems of limited bandwidth, device diversity, unreliable connectivity, latency, and limited energy are some of the issues that distributed learning in wireless networks needs to overcome. This will directly impact the ability to deploy EI at the edge, as the generation of explanations and transferring the model to the edge will require extra resources in terms of computation and communication. It is important, therefore, to design secure and explainable edge frameworks with consideration of constraints at the network level and performance of the AI [27].

For privacy-preserving AI, a new architectural approach that combines cloud, edge, and end devices is emerging as a viable option: federated cloud–edge–end collaboration. The lightweight, safe federated learning architectures allow computational tasks to be distributed amongst cloud servers, edge nodes and end devices, while retaining privacy and scalability. These architectures are especially well-suited for environments where computational resources are limited, and updates, aggregation and inference of the model can be distributed across various layers of computation depending on the security and resource restrictions [28]. The multi-layered partnership enhances the ability to implement reliable AI in distributed systems. Preserving data security and privacy are always important issues with edge-IoT federated learning. Data leakage, poisoning attack, inference attack, unauthorized access and compromised devices are all potential dangers in the edge environment. While federated learning can minimize the direct contact of raw data, it is not free from security risks produced by model updates, gradients and aggregations [29]. Edge intelligence frameworks need to be robust, secure, and include encryption, authentication, anomaly detection, privacy-preserving aggregation and strong model validation mechanisms.

Blockchain technologies can be used to further improve the process of making secure and explainable decisions by fostering decentralized trust, auditability and tamper-resistance of the record of AI-driven processes. With explainable AI, blockchain can be used to keep a record of decision trails, the validation of model results, and enhance the accountability of actions taken by distributed actors in the detection of cyber threats, fostering transparency. With explainable AI, blockchain can be used to keep a record of decision trails, validate AI model outputs, and ensure accountability of actions taken by distributed actors in detecting cyber threats, providing transparency [30]. The integration is a boon in environments where several stakeholders are in interaction without any trusted authority. It's also vital to have a comparative overview of the differences between AI, machine learning and deep learning to help determine the appropriate edge intelligence framework. The complexity, interpretability, computational cost and applicability of these model families vary [31]. Therefore, the choice of models along with the deployment architecture, privacy mechanism and explanation approach should be consistent with the operation needs of smart digital ecosystems and securely explained. The architecture of secure and explainable edge intelligence is depicted in Figure 1, where the various components of the IoT, the edge computing resources, the techniques for explainable AI, federated learning, cloud services and the security mechanisms work together to enable secure and trustworthy decision-making in smart digital ecosystems.

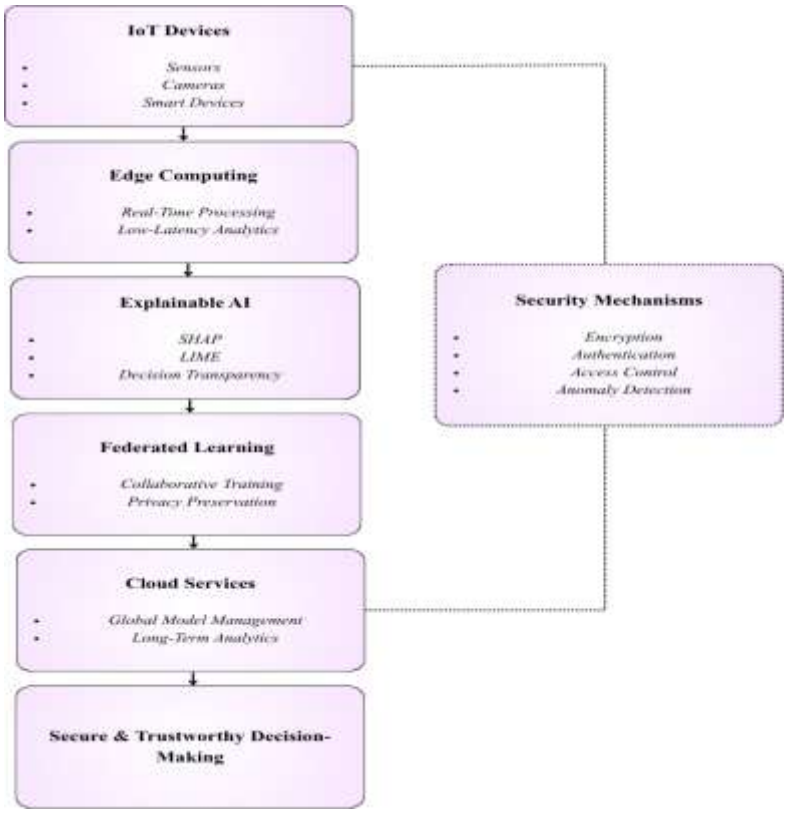

Figure 1. Architecture of Secure and Explainable Edge Intelligence in Smart Digital Ecosystems

6. Security-Aware Explainable Decision-Making Mechanisms

To keep intelligent systems in smart digital ecosystems trustworthy, resilient and accountable, security-aware explainable decision-making mechanisms are essential. The attack surface of IoT devices skyrockets as digital ecosystems grow and spread over interconnected platforms, IoT devices, cloud infrastructures, edge nodes and autonomous services. These settings are hard to secure due to the presence of multiple threats stemming from technical vulnerabilities, lack of good governance, data disintegration, and interdependencies between multiple parties [32]. So not only should decision-making frameworks identify and react to threats, but they should also provide transparency of why certain security actions are being taken. One of the serious concerns of using machine learning in cybersecurity systems is adversarial machine learning. Attackers might be able to alter the input data, poison training sets, exploit model vulnerabilities or create adversarial examples that will trick the detection system. Such threats are especially hazardous in edge-based environments where some devices in the network might exist with restricted resources and centralized supervision. The successful deployment of effective defense mechanisms should involve a strong model design, adversarial training, anomaly detection, and continuous monitoring, as well as the ethical issues involved in automated cyber decisions [33].

Explainable AI (XAI) is a key component in enhancing the trust of autonomous cyber decision infrastructures. Security analysts and system operators need to understand why an automated alert was generated, why the risk score is so high, why access control is set, what the recommended response is, and more. Technically correct answers may also be given by autonomous cyber systems without an explanation, which will be hard to validate, challenge, and act upon. XAI can help with this by providing transparency and security operations, enhancing human oversight and accountability, and building institutional trust in human-robot interactions [34]. Explainable intrusion detection is emerging as one of the key application fields of security-aware XAI. IoT ecosystems produce massive amounts of diverse types of traffic data, and traditional intrusion detection systems are becoming more challenging. XAI-based IDSs are able to explain the features, traffic or context that led to a detection result and to detect malicious behavior. This enhances the interpretability of analysts, ambiguity of false alarms is reduced, and it will enable more effective cyber defense decisions in distributed IoT settings [35].

Trustworthiness of distributed AI systems should consider aspects such as robustness, privacy, and governance. Potential problems of drift and malicious behavior, privacy leakage, biases, and unreliable aggregation could arise due to the fact that the distributed edge environment includes training, retraining, and deployment of AI models across different nodes. Security in learning protocols, privacy-preserving computation, strong validation, and transparent auditing and governance institutions that create accountability throughout the AI lifecycle are needed to ensure the trustworthiness of the AI mechanisms [36]. Additional regulation and governance also impact the deployment of explainable and secure AI systems. The risk of the various application scenarios is different, and AI governance frameworks need to assess systems based on their operational circumstances, possible risks, and accountability needs, as well as on compliance. In this governance process, security-aware explainability can serve as a tool that provides traceable evidence of the action of the model(s) and the rationale behind the decisions they make, as well as the risk management practices used in the model(s) [37]. Explainable decision-making mechanisms that are security-enabled should be integrated in various stages of smart digital ecosystem design, deployment and monitoring. Adversarial robustness, explainable intrusion detection, privacy-preserving distributed AI, and accountability-driven by governance principles offer a base for resilient intelligent systems. This is particularly crucial at the edge, where decisions need to be made quickly, locally, securely and yet comprehensible to humans involved in the decision-making process. Figure 2 shows the flow of security-aware explainable decision-making, which involves identifying cyber threats using AI-based detection processes, interpreting the threats with explainable AI, and converting the threat into security responses that increase trust, accountability and governance.

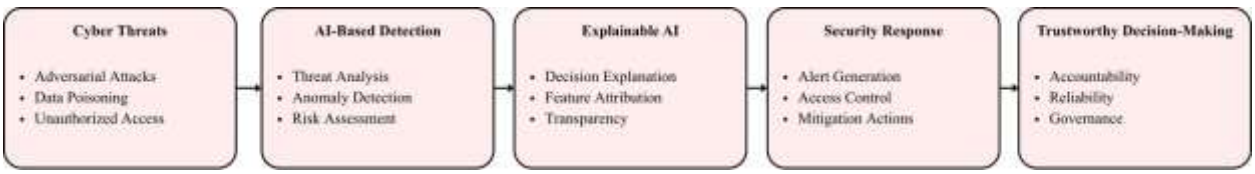

Figure 2. Security-Aware Explainable Decision-Making Workflow

7. Application Domains and Real-World Implementations

As data-driven decisions are made in a real-world, smart digital ecosystem, AI with explanations and edge-based decisions are becoming more relevant than ever to ensure accuracy, transparency, security, and context sensitivity of decisions. One of the most important application areas is healthcare, where AI systems can facilitate various aspects of diagnosis, treatment planning, risk prediction, and optimizing the clinical workflow. With the addition of explainability within the black box AI algorithms used within health care, it has become easier for physicians to understand how to interpret the results generated by AI, the accuracy of the AI algorithm, and the usefulness of the AI in helping them make their decision [38]. This is because besides being technically sound, the decisions made in medicine have to be humanly accountable, traceable, and trustworthy. For some medical specialties, it is also vital that the AI system should have the ability of uncertainty awareness. For musculoskeletal medicine, quantification of uncertainty is critical because the predictions made by AI can influence diagnosis, surgery, rehabilitation and risk assessment. Besides the ability to produce a classification, or a recommendation, in this context a system should be able to communicate confidence levels, limitations and decisions to be uncertain or unsure [39]. The need for XAI in medical imaging adds to this. Techniques used to retake images for the shoulder joint based on deep learning can be explained by XAI and can help the user understand the reasons

for retaking the image and making automatic decisions regarding quality control for the image [40]. As the examples illustrate, explainability in the context of healthcare AI should be not only clinically relevant for the predictions but should also increase the accuracy of the prediction.

Secure and explainable intelligence is another important application area: Smart Cities. The transportation, surveillance, energy, emergency management, public services and environmental monitoring systems are important and interconnected infrastructures in urban ecosystems. The need for a secure and reliable XAI framework for smart city applications is to enhance transparency in automated decisions, mitigate risks such as privacy violations, cyberattacks, and dependence on infrastructure [41]. Edge computing can be especially useful in these environments, where many decisions need to be made near the data-generating sources (e.g. cameras, traffic sensors, public safety systems) where time-sensitive and reliable requirements are paramount. Another change is in the industrial space towards human-centric intelligent systems. Industry 5.0 puts humans in the center of the picture, with a focus on how collaboration between humans and machines and intelligent digital infrastructure contributes to explainability, trust and worker-centered decision support. Human-centric AI architectures for Industry 5.0 demonstrate the benefits of explainable and adaptive systems in improving manufacturing and maintenance, quality control, and human-machine collaboration [42]. Explanations must not only be based on the outputs of the models, but rather relate to what the engineers, operators and managers can do in this context. Explainable AI can also be used to enable trustworthy decision-making in other areas, such as smart energy and sustainable infrastructure. The need for infrastructure to be energy efficient is associated with the need for measurement, verification and optimization to achieve a net zero carbon. Explainability of effective AI solutions can enhance trust in energy performance assessment and provide insights into energy savings calculations, anomalies, or efficiency gains [43]. This is especially important due to the nature of some sustainability decisions, which can be long-term investments, regulatory reporting and multi-stakeholder responsibility. Secure and Explainable Mechanisms are needed for Cyber-physical Systems as they involve digital intelligence in the operations of the physical world. In the context of cyber-physical systems (CyPS), hybrid XAI systems for adversarial attack detection illustrate the use of explainability to strengthen security by uncovering patterns in adversarial attacks, vulnerabilities in the models, and unusual system behavior [44]. Throughout the healthcare, smart cities, Industry 5.0, energy infrastructure and cyber-physical systems sectors, the real-world experience demonstrates the need for explainable edge intelligence that is domain-aware, security-aware and actionable. Table 3 summarizes the key application scenarios and the extent to which explainable edge intelligence can help to make decisions in a secure and trustworthy way.

Table 3. Application Domains of Secure Explainable Edge Intelligence

Application Domain	Decision Context	XAI Contribution	Edge Computing Role	Security Requirement
Healthcare	Diagnosis, imaging, clinical support	Improves clinician trust and interpretation	Enables local and timely analysis	Privacy and safety assurance
Smart Cities	Traffic, surveillance, public services	Explains automated urban decisions	Supports real-time sensor processing	Secure infrastructure operation
Industry 5.0	Maintenance, quality control, human-machine collaboration	Supports human-centric decisions	Enables factory-level intelligence	Protection of industrial assets
Smart Energy	Efficiency measurement and infrastructure monitoring	Clarifies energy predictions and anomalies	Enables distributed energy analytics	Integrity of energy data
Cyber-Physical Systems	Attack detection and autonomous control	Reveals abnormal behavior and attack patterns	Supports rapid local response	Adversarial robustness
IoT Ecosystems	Distributed sensing and automation	Explains device-level decisions	Reduces cloud dependency	Secure communication and access control

8. Open Challenges and Future Research Directions

While there have been considerable advances, explainable AI in the context of secure edge-based decision making remains with many technical and practical challenges. Existing XAI techniques range from assuming a specific objective to providing differing explanations, requiring varying computational resources and being appropriate for differing situations of real-time deployment. There are many approaches that are not easily transferable between domains, as explanations that fit into one may not fit into the other. While there are existing surveys that point to the need for better explanation fidelity, stability, scalability, and user interpretability, there is a need for more standardized and context-aware XAI frameworks [45]. The integration of robustness, fairness, privacy, accountability, and human oversight is also a critical strength to build into trustworthy explainable AI. Recent developments have brought models and algorithms closer to the human understanding of their underlying principles, but many models are still not technically accurate and comprehensible to users who are not experts. Moving beyond post-hoc visualization would be beneficial for future research; moving to trustworthy explanation pipelines that can be validated, audited and modified to changing environments would be beneficial for future research [46]. For edge computing environments, models can be subject to limited resources and changing data conditions.

The challenge of implementing XAI in high-risk settings is exemplified by the deployment of clinical decision support systems. The deployment of clinical decision support systems is a good example of the wider issue of implementing XAI in high-risk settings. In health care, explanations need to be designed to facilitate the comprehension of clinicians, the safety of patients, and the ethical accountability of both patients and clinicians, as well as compliance with regulatory requirements. But current XAI approaches are often not easily integrated into clinical workflow, domain knowledge, and clinical decision making [47]. Therefore, future studies should highlight the design of explanations that consider humans, domain-specific validation, and collaborative decision-making between humans and AI systems. One of the most significant challenges that still exists in the field of XAI is evaluation and validation. If there are no effective means of evaluating, it is hard to judge whether explanations are correct, helpful, convincing, and/or misleading. Benchmarks, quantitative and qualitative assessment metrics and validation procedures for edge-based explainable systems should be developed in the future. This is especially important for autonomous and cyber-physical systems, where misunderstandings or misinterpretations in explanations can result in wrong or unsafe decisions [48].

One more research direction is XAI for edge devices, which is the development of resource-efficient XAI for edge devices. Some of the explanation techniques require significant computational, memory, or communication bandwidth, making them hard to deploy on the constrained IoT nodes and embedded platforms. Lightweight explanation models, adaptive explanation generation, model compression, neuromorphic computing and hardware-aware XAI are future frameworks to investigate. These techniques can contribute to generating explanations without impacting latency, energy costs, or security. Further integration of explainable intelligence with Federated Learning, Blockchain, Digital Twins, 5G/6G networks and Privacy-preserving computation is necessary. Future systems should enable explanations throughout distributed learning processes, detect unauthorized and biased model updates and offer security-sensitive explanations of audit trails. Explainability should also be integrated into governance structures so that organizations can track the behavior of their models, provide accountability and ensure compliance.

Future studies are needed in the direction of secure explainable edge intelligence frameworks rather than individual XAI tools for AI. These frameworks need to be technically correct, explained well, secure, resilient to threats, and user-friendly while respecting privacy. This integration will be crucial for developing smart digital ecosystems that are trusted and can inform critical decisions across a range of sectors, from healthcare, smart cities and Industry 5.0, energy systems, to cyber-physical infrastructures.

9. CONCLUSION

In the discussion, the intersection of the concepts of Explainable AI, Edge Computing and Cybersecurity in the context of Secure Decision Making for Smart Digital Ecosystems was discussed. It becomes clear from the review of the literature that the future belongs to the decision making architectures which should be precise and prompt, as well as capable of providing transparency, privacy, and cybersecurity. It gives rise to the emergence of greater real-time intelligence through edge computing and enhanced AI decision interpretability and accountability through XAI. These technologies, together with other promising ones like federated learning, blockchain, intrusion detection and governance-centric security solutions, form a powerful base for development of trustworthy distributed intelligence solutions. At the same time, the review reveals possible areas of the application of the technologies under consideration in which the positive impact on the human trust and reliability would be achieved. Those areas include healthcare, smart cities, Industry 5.0, smart energy, IoT ecosystem, and cyber-physical systems. Still, there remain problems related to the development of the resource-efficient explainability, validation framework, adversarial robustness, privacy protection, and context-aware governance. In order to address them, the future research should focus on creation of the unified framework for the integration of explainability, security, edge intelligence, and human-centered design in the development stage of the intelligent ecosystem.

REFERENCES

1. Ocampo A, Antonino PO, Capilla R, Nakagawa EY. Digitalization of smart ecosystems. IEEE Software. 2025 Feb 13;42(2):33-7.
2. Komninos N. Smart cities and connected intelligence: Platforms, ecosystems and network effects. Routledge; 2019 Dec 5.
3. Narne S, Adedoja T, Mohan M, Ayyalasomayajula T. AI-driven decision support systems in management: enhancing strategic planning and execution. International journal on recent and innovation trends in computing and communication. 2024 Mar 16;12(1):268-76.
4. Ravi M, Negi A, Bommi NS, Rouf N. Evolution of AI-driven decision making with decision support systems, expert systems, recommender systems, and XAI. IETE Technical Review. 2025 Jul 4;42(4):428-65.
5. Zhou Z, Chen X, Li E, Zeng L, Luo K, Zhang J. Edge intelligence: Paving the last mile of artificial intelligence with edge computing. Proceedings of the IEEE. 2019 Jun 12;107(8):1738-62.
6. Konda PR. AI-DRIVEN CLOUD DATA ANALYTICS FRAMEWORK FOR INTELLIGENT ENTERPRISE DECISION SYSTEMS. Indonasian Journal of Advanced Research & Technology. 2024 Nov 21;6(6).
7. Cioroaica E, Kuhn T, Buhnova B. (Do not) trust in ecosystems. In2019 IEEE/ACM 41st International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER) 2019 May 25 (pp. 9-12). IEEE.
8. Qudus L. Advancing cybersecurity: strategies for mitigating threats in evolving digital and IoT ecosystems. Int Res J Mod Eng Technol Sci. 2025 Jan;7(1):3185.
9. Katuk N, Abdullah WA, Sugiharto T, Ahmad I. Smart technology: Ecosystem, impacts, challenges and the path forward. Information System and Smart City. 2023 Nov 9;3(1):63-.

10. Nativi S, Mazzetti P, Craglia M. Digital ecosystems for developing digital twins of the earth: The destination earth case. Remote Sensing. 2021 May 28;13(11):2119.
11. Petrakis K, Agorogiannis E, Antonopoulos G, Anagnostopoulos T, Grigoropoulos N, Veroni E, Berne A, Azaiez S, Benomar Z, Kakoulidis H, Prasinos M. Enhancing DevOps practices in the IoT–Edge–Cloud continuum: Architecture, integration, and software orchestration demonstrated in the COGNIFOG framework. Software. 2025 Apr 15;4(2):10.
12. Chatterjee B, Cao N, Raychowdhury A, Sen S. Context-aware intelligence in resource-constrained IoT nodes: Opportunities and challenges. IEEE Design & Test. 2019 Feb 13;36(2):7-40.
13. Angelov PP, Soares EA, Jiang R, Arnold NI, Atkinson PM. Explainable artificial intelligence: an analytical review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2021 Sep;11(5):e1424.
14. Confalonieri R, Coba L, Wagner B, Besold TR. A historical perspective of explainable artificial intelligence. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2021 Jan;11(1):e1391.
15. Oluoha OM, Odeshina AB, Reis OL, Okpeke FR, Attipoe VE, Orieno O. A privacy-first framework for data protection and compliance assurance in digital ecosystems. Iconic Research and Engineering Journals. 2023 Oct;7(4):620-46.
16. 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 Nov 9;14(1):1-59.
17. Ofili BT, Obasuyi OT, Akano TD. Edge Computing, 5G, and Cloud Security Convergence: Strengthening USA's Critical Infrastructure Resilience. Int J Comput Appl Technol Res. 2023;12(9):17-31.
18. Vermesan O, Piuri V, Scotti F, Genovese A, Labati RD, Coscia P. Explainability and interpretability concepts for edge ai systems. InAdvancing edge artificial intelligence 2024 Mar 4 (pp. 197-227). River Publishers.
19. Puthanveettil Madathil A, Luo X, Liu Q, Walker C, Madarkar R, Cai Y, Liu Z, Chang W, Qin Y. Intrinsic and post-hoc XAI approaches for fingerprint identification and response prediction in smart manufacturing processes. Journal of Intelligent Manufacturing. 2024 Dec;35(8):4159-80.
20. Chinnaraju A. Explainable AI (XAI) for trustworthy and transparent decision-making: A theoretical framework for AI interpretability. World Journal of Advanced Engineering Technology and Sciences. 2025 Mar 30;14(3):170-207.
21. Presciuttini A, Cantini A, Portioli-Staudacher A. From explanations to actions: leveraging SHAP, LIME, and counterfactual analysis for operational excellence in maintenance decisions. In2024 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) 2024 Nov 4 (pp. 1-6). IEEE.
22. Shiri FM, Perumal T, Mustapha N, Mohamed R. A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU. arXiv preprint arXiv:2305.17473. 2023 May 27.
23. Holzinger A, Saranti A, Hauschild AC, Beinecke J, Heider D, Roettger R, Mueller H, Baumbach J, Pfeifer B. Human-in-the-loop integration with domain-knowledge graphs for explainable federated deep learning. InInternational Cross-Domain Conference for Machine Learning and Knowledge Extraction 2023 Aug 22 (pp. 45-64). Cham: Springer Nature Switzerland.
24. Assis A, Dantas J, Andrade E. The performance-interpretability trade-off: a comparative study of machine learning models. Journal of Reliable Intelligent Environments. 2025 Mar;11(1):1.
25. Jagatheesaperumal SK, Pham QV, Ruby R, Yang Z, Xu C, Zhang Z. Explainable AI over the Internet of Things (IoT): Overview, state-of-the-art and future directions. IEEE Open Journal of the Communications Society. 2022 Oct 26;3:2106-36.
26. Liu J, Huang J, Zhou Y, Li X, Ji S, Xiong H, Dou D. From distributed machine learning to federated learning: A survey. arXiv preprint arXiv:2104.14362. 2021 Apr 29.
27. Chen M, Gündüz D, Huang K, Saad W, Bennis M, Feljan AV, Poor HV. Distributed learning in wireless networks: Recent progress and future challenges. IEEE journal on selected areas in communications. 2021 Oct 6;39(12):3579-605.
28. Zhan S, Huang L, Luo G, Zheng S, Gao Z, Chao HC. A review on federated learning architectures for privacy-preserving AI: Lightweight and secure cloud–edge–end collaboration. Electronics. 2025 Jun 20;14(13):2512.
29. Li H, Ge L, Tian L. Survey: federated learning data security and privacy-preserving in edge-Internet of Things. Artificial Intelligence Review. 2024 Apr 29;57(5):130.
30. Kumar P, Javeed D, Kumar R, Islam AN. Blockchain and explainable AI for enhanced decision making in cyber threat detection. Software: Practice and Experience. 2024 Aug;54(8):1337-60.
31. Lamaazi H, Mathew E. Comprehensive comparative analysis of artificial intelligence, machine learning, and deep learning. InArtificial Intelligence for Intelligent Systems 2024 Jul 31 (pp. 51-68). CRC Press.
32. Tabar S, Towhidi G, Bryant T. Overview of the challenges and potential solutions in securing digital ecosystem in the US. Issues in Information Systems. 2025 Jan 1;26(1):79-93.
33. Singh V, Singh S. Adversarial Machine Learning in Cybersecurity: Threat Detection, Defense Mechanisms, and Ethical Issues. InChallenges and Solutions for Cybersecurity and Adversarial Machine Learning 2025 (pp. 103-136). IGI Global Scientific Publishing.
34. KM Z, Akhtaruzzaman K, Tanvir Rahman A. Building trust in autonomous cyber decision infrastructure through explainable AI. International Journal of Economy and Innovation. 2022 Nov 29;29:405-28.
35. Moustafa N, Koroniotis N, Keshk M, Zomaya AY, Tari Z. Explainable intrusion detection for cyber defences in the internet of things: Opportunities and solutions. IEEE Communications Surveys & Tutorials. 2023 May 26;25(3):1775-807.
36. Wei W, Liu L. Trustworthy distributed ai systems: Robustness, privacy, and governance. ACM Computing Surveys. 2025 Feb 10;57(6):1-42.
37. Szadeczky T, Bederna Z. Risk, regulation, and governance: evaluating artificial intelligence across diverse application scenarios: Risk, regulation, and governance. Security Journal. 2025 Mar;38(1):35.

38. Srinivasu PN, Sandhya N, Jhaveri RH, Raut R. From blackbox to explainable AI in healthcare: existing tools and case studies. Mobile Information Systems. 2022;2022(1):8167821.
39. Vahdani AM, Shariatnia M, Rajpurkar P, Pareek A. Towards trustworthy artificial intelligence in musculoskeletal medicine: A narrative review on uncertainty quantification. Knee Surgery, Sports Traumatology, Arthroscopy. 2025 Sep;33(9):3418-37.
40. Sekiura K, Yoshimura T, Sugimori H. Development of an XAI-Enhanced Deep-Learning Algorithm for Automated Decision-Making on Shoulder-Joint X-Ray Retaking. Applied Sciences. 2025 Sep 29;15(19):10534.
41. Al Garni M, Mishra S. A secure and reliable framework for explainable artificial intelligence (XAI) in smart city applications.
42. Rožanec JM, Novalija I, Zajec P, Kenda K, Tavakoli Ghinani H, Suh S, Bian S, Veliou E, Papamartzivanos D, Giannetsos T, Menesidou SA. Human-centric artificial intelligence architecture for industry 5.0 applications. International journal of production research. 2023 Oct 18;61(20):6847-72.
43. Moraliyage H, Dahanayake S, De Silva D, Mills N, Rathnayaka P, Nguyen S, Alahakoon D, Jennings A. A robust artificial intelligence approach with explainability for measurement and verification of energy efficient infrastructure for net zero carbon emissions. Sensors. 2022 Dec 5;22(23):9503.
44. Taufik M, Aziz MS, Fitriana A. Hybrid Explainable AI (XAI) Framework for Detecting Adversarial Attacks in Cyber-Physical Systems. Journal of Technology Informatics and Engineering. 2025 Apr 22;4(1):157-71.
45. Yang W, Wei Y, Wei H, Chen Y, Huang G, Li X, Li R, Yao N, Wang X, Gu X, Amin MB. Survey on explainable AI: From approaches, limitations and applications aspects. Human-Centric Intelligent Systems. 2023 Sep;3(3):161-88.
46. Rawal A, McCoy J, Rawat DB, Sadler BM, Amant RS. Recent advances in trustworthy explainable artificial intelligence: Status, challenges, and perspectives. IEEE Transactions on Artificial Intelligence. 2021 Dec 10;3(6):852-66.
47. Antoniadi AM, Du Y, Guendouz Y, Wei L, Mazo C, Becker BA, Mooney C. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. Applied Sciences. 2021 May 31;11(11):5088.
48. Mahto MK. Explainable artificial intelligence: Fundamentals, approaches, challenges, XAI evaluation, and validation. InExplainable artificial intelligence for autonomous vehicles 2024 Aug 14 (pp. 25-49). CRC Press.