UDC 667.021.1 DOI https://doi.org/10.35546/kntu2078-4481.2025.3.2.19

O. Y. HONSOR
Candidate of Technical Sciences,
Associate Professor at the Department of Specialized Computer Systems
National University "Lviv Polytechnic"
ORCID: 0000-0003-0895-5859

TRUSTWORTHY AI IN SMART SENSOR SYSTEMS IN IOT: CHALLENGES AND SOLUTIONS

This article discusses the key challenges and prospects for implementing reliable artificial intelligence (AI) in intelligent sensor systems of the Internet of Things (IoT), with a special focus on the principles of trust, security, explainability, and legal compliance. The integration of AI with sensor networks allows for autonomous, context-sensitive real-time decision-making, which is critical for healthcare, industry, transportation, and environmental monitoring.

The paper presents an interdisciplinary review of current scientific approaches, including the use of Federated Learning, Explainable AI, secure data processing methods at the edge, Edge Intelligence, as well as risk assessment and ethical compliance in accordance with international standards such as GDPR, ISO/IEC 42001, NIST AI RMF, IEEE 7000, and others. The analysis also covers the issues of data quality and integrity, security, transparency, bias, reliability, and regulatory uncertainties that arise when implementing AI in IoT systems.

The author proposes a conceptual model of trust in AI based on seven interrelated factors: data quality, reliability, security, privacy protection, explainability, fairness, and legal compliance. The model allows to quantify the level of trust in the system by adapting the weighting factors to the specifics of a particular application.

In addition to analyzing the problems, the authors provide practical recommendations for overcoming challenges, such as implementing real-time data quality checks, building a secure architecture based on the secure-by-design principle, using XAI algorithms (SHAP, LIME), and conducting fairness and compliance audits. The table of correspondences between challenges and solutions allows a systematic approach to the implementation of trusted AI in complex distributed systems.

Thus, this work forms the basis for the development of policies, standards, and strategies for the implementation of trusted artificial intelligence in IoT environments, maintaining a balance between technological innovation and ethical and legal responsibility.

Key words: *Trustworthy AI, Smart Sensor Systems, Internet of Things, Explainable AI.*

О. Й. ГОНСЬОР
кандидат технічних наук,
доцент кафедри спеціалізованих комп'ютерних систем
Національний університет «Львівська політехніка»
ORCID: 0000-0003-0895-5859

ДОСТОВІРНИЙ ШІ В ІНТЕЛЕКТУАЛЬНИХ СЕНСОРНИХ СИСТЕМАХ IOT: ВИКЛИКИ ТА РІШЕННЯ

У статті розглядаються ключові виклики та перспективи впровадження надійного штучного інтелекту (AI) в інтелектуальні сенсорні системи Інтернету речей (IoT), з особливим акцентом на принципи довіри, безпеки, пояснюваності та правової відповідності. Інтеграція AI з сенсорними мережами дозволяє досягати автономного, контекстно-чутливого прийняття рішень у реальному часі, що є критично важливим для галузей охорони здоров'я, промисловості, транспорту та екологічного моніторингу.

В роботі представлено міждисциплінарний огляд сучасних наукових підходів, включаючи використання федеративного навчання (Federated Learning), пояснюваного AI (XAI), методів безпечної обробки даних на периферії (Edge Intelligence), а також оцінку ризиків і забезпечення етичної відповідності згідно з міжнародними стандартами – такими як GDPR, ISO/IEC 42001, NIST AI RMF, IEEE 7000 та інші. Проведений аналіз також охоплює проблеми якості та цілісності даних, безпеки, прозорості, упередженості, надійності, а також нормативно-правових невизначеностей, які виникають при впровадженні AI в IoT-системи.

Запропоновано концептуальну модель довіри до AI, яка базується на семи взаємопов'язаних факторах: якість даних, надійність, безпека, захист конфіденційності, пояснюваність, справедливість і правова відповідність. Модель дозволяє кількісно оцінити рівень довіри до системи, адаптуючи вагові коефіцієнти під специфіку конкретного застосування.

Окрім аналізу проблем, автори надають практичні рекомендації для подолання викликів, таких як впровадження перевірки якості даних у реальному часі, побудова безпечної архітектури за принципом «безпека

за задумом» (secure-by-design), застосування XAI-алгоритмів (SHAP, LIME), проведення аудитів справедливості та відповідності нормативним актам. Таблиця відповідностей між викликами і рішеннями дозволяє системно підійти до реалізації довіреного AI у складних розподілених системах.

Таким чином, дана робота формує основу для формування політик, стандартів та стратегій впровадження довіреного штучного інтелекту в середовищах Інтернету речей, підтримуючи баланс між технологічною інновацією та етично-правовою відповідальністю.

Ключові слова: *Достовірний ШІ, Розумні Сенсорні Системи, Інтернет Речей, Пояснюваний ШІ.*

Problem statement

Smart sensor systems are intelligent devices that collect, process, and transmit data within the Internet of Things (IoT) ecosystem. These sensors integrate artificial intelligence (AI), machine learning (ML), and edge computing to enhance automation, efficiency, and real-time decision-making. In the context of Industry 4.0, these systems play a pivotal role in predictive maintenance, quality control, and supply chain optimisation by facilitating autonomous operations.

AI improves sensor networks by analysing vast amounts of data, identifying patterns and making autonomous decisions with minimal human intervention. However, an over-reliance on artificial intelligence can lead to bias and reduced transparency and accountability in decision-making.

Trust is imperative for successfully integrating artificial intelligence (AI) within the industrial Internet of Things (IoT) domain. Ensuring that artificial intelligence (AI) aligns with regulatory imperatives, including those stipulated by the General Data Protection Regulation (GDPR), the International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 42001 standards, and the EU's AI Act, is of paramount importance. The development of reliable AI hinges upon establishing governance frameworks that strike a balance between the pursuit of innovation and the implementation of legal and ethical safeguards.

The objective of this article is to furnish policy and legal strategy recommendations that aim to fortify trust, enhance accountability, and bolster compliance in the context of AI-driven IoT applications.

Analysis of Recent Research and Publications

Recent advances in integrating artificial intelligence with intelligent sensor systems within the Internet of Things have been accompanied by significant research aimed at ensuring the reliability of these systems. A review of the seminal scientific contributions from recent years reveals a concerted effort to address the challenges and propose solutions that guarantee the reliability of artificial intelligence in intelligent sensor systems.

Federated Learning (FL) has emerged as a pivotal strategy for enhancing privacy and security in distributed AI systems. In [1], FederatedTrust is presented, an algorithm designed to calculate the trustworthiness of FL models by evaluating six pillars: privacy, reliability, fairness, explainability, accountability, and federation. Their work underscores the necessity for comprehensive metrics to assess and ensure the reliability of AI models in decentralised environments.

Authors in [2, 3] conducted a survey on explainable artificial intelligence (XAI) techniques tailored for the IoT, discussing their applicability in areas such as security enhancement and the Industrial Internet of Things IIoT. The authors advocate for integrating XAI to enhance transparency and user trust in AI-driven IoT systems.

In [4], a comprehensive guide is provided for developing trustworthy AI systems, addressing aspects such as robustness, fairness, privacy preservation, and accountability. The researchers propose a systematic approach that spans the entire AI lifecycle, offering actionable insights for practitioners to enhance trustworthiness in AI applications.

The convergence of Edge Computing and AI, known as Edge Intelligence (EI), gives rise to resource constraints and security challenges. In their seminal work, Wang et al. [5] presented a comprehensive survey on the subject of trustworthy EI, delineating its fundamental characteristics and a multi-layered architectural framework. They discuss enabling technologies and solutions aimed at achieving trustworthiness in EI systems, emphasising the importance of security, reliability, transparency, and sustainability.

Integrating artificial intelligence (AI) within the Internet of Things (IoT) systems introduces significant cybersecurity risks. A recent article in Frontiers in Big Data [6–8] discusses the importance of explainability and transparency in AI/ML models for dependency risk analysis in IoT. The authors of this article emphasise the importance of explainable artificial intelligence (XAI) techniques, interpretable models, and human-in-the-loop systems to enhance trustworthiness in AI-driven IoT applications.

In [9–11], authors examined the application of artificial intelligence (AI) and the Internet of Things (IoT) technologies in the context of environmental pollution monitoring and management. The authors expound on the intricacies of predicting and tracking alterations in pollution levels, underscoring the promise of AI-driven sensor systems in confronting environmental predicaments. They further emphasise the necessity of reliable AI in these applications, underscoring the need for rigorous research and development to ensure the integrity and efficacy of these technologies.

A collective examination of these studies underscores the multifaceted efforts underway to ensure that AI-integrated smart sensor systems in IoT are developed and deployed in a trustworthy manner. Such efforts address challenges related to privacy, security, explainability, and sustainability.

Formulation of Research Objective

Integrating artificial intelligence with intelligent sensor systems enables autonomous, context-sensitive, real-time decision-making in various industries, including healthcare, manufacturing, transportation, and environmental monitoring. However, these advancements have given rise to significant concerns regarding the reliability of such systems. The concept of trustworthiness in the context under consideration encompasses several interrelated attributes.

- **Reliability:** The AI system must function accurately and consistently under various operational conditions.
- **Transparency and Explainability:** Stakeholders must understand how AI models make decisions, especially in safety-critical or regulated environments.
- **Security and Privacy:** As sensor systems collect and process sensitive data, ensuring data integrity and privacy is crucial. Techniques like federated learning, secure multiparty computation, and differential privacy are increasingly employed.
- **Ethical and Legal Compliance:** Trustworthy AI systems must adhere to ethical principles and comply with legal standards such as the GDPR, AI Act, and sector-specific guidelines.
- **Accountability and Fairness:** The AI's decision-making process must be auditable, and outcomes must not reflect systemic bias or discrimination.

Recent literature has proposed various strategies to enhance trust, such as adopting Explainable AI (XAI) models, AI risk assessment frameworks, and standardisation efforts (e.g., IEEE P7000 series, ISO/IEC 42001). Moreover, trustworthiness is increasingly treated as a multidisciplinary challenge involving not only technical robustness but also human-centric and governance perspectives.

Table 1 presents a compendium of pivotal standards and regulations pertaining to trustworthiness in AI-driven smart sensor systems, methodically organised by category.

These standards encompass critical domains such as AI ethics, safety, cybersecurity, data protection, and system reliability, which are paramount in IoT environments.

Table 1

Standards and regulations pertaining to trustworthiness in AI-driven smart sensor systems

Category	Standard / Regulation	Scope / Description	Relevance to Trustworthiness
AI Governance	EU AI Act (2024/25)	Risk-based regulation of AI systems in the EU	Requires transparency, risk management, and oversight
	NIST AI RMF (2023)	US framework for managing AI risks	Emphasises reliability, explainability, and fairness
	ISO/IEC 42001 (2023)	AI Management System Standard	Operationalises trustworthy AI in organisations
	OECD AI Principles (2019)	Global ethical guidelines for AI development	Promotes human-centric, fair, and transparent AI
Ethics & Explainability	IEEE 7000 Series	Standards for the ethical design of autonomous systems	Covers transparency, bias mitigation, and privacy
	IEEE 7001, 7002, 7003	Focus on transparency, privacy, and algorithmic fairness	Ensures ethical deployment of AI in sensor networks
Functional Safety	IEC 61508	Functional safety for electronic systems	Ensures fail-safe design of AI-integrated sensor systems
	ISO 13849	Safety of machinery with embedded control systems	Used in industrial AI-driven automation
	ISO 26262	Functional safety in automotive systems	Applies to AI with smart sensors in autonomous vehicles
Cybersecurity	IEC 62443	Cybersecurity for industrial control systems	Protects sensor-based AI in manufacturing
	ISO/IEC 27001	Information security management systems	Ensures confidentiality and data integrity
	NIST SP 800-53/800-82	Security for federal and industrial control systems	Provides cybersecurity guidelines for smart IoT systems
Privacy & Data	GDPR (EU)	Regulation on data protection and privacy	Requires consent, transparency, and explanation rights
	CCPA / CPRA (US)	US data privacy regulations for consumers	Controls handling of personal sensor data
	ISO/IEC 27701	Privacy Information Management System	Extension of ISO 27001 for managing PII
System Quality	ISO/IEC 25010	Software/system quality models	Ensures reliability, usability, and maintainability
	ISO 9001	General quality management systems	Supports consistent and safe development practices

Presentation of the Main Research Material Trust Challenges in AI-based IoT

Integrating artificial intelligence (AI) into Internet of Things (IoT) ecosystems, particularly those that incorporate intelligent sensor systems, has introduced a novel dimension of intelligence and autonomy. These systems have found widespread application in domains such as predictive maintenance in manufacturing, real-time health monitoring, and the management of smart city infrastructure. Concurrently, implementing AI in sensor systems introduces specific challenges related to data trustworthiness and the reliability and ethical use of AI models. Addressing these challenges is imperative to ensure safe, dependable, and ethical operations.

Data Quality and Integrity: The primary concern pertains to the integrity and reliability of the data. The integrity of the data received from sensors is paramount to operating artificial intelligence models, which rely on this data to make accurate predictions or decisions. However, IoT environments are inherently dynamic and subject to various physical and digital stresses. Drift, calibration errors, hardware failures, or data transmission delays can cause data streams to become corrupted or biased. This compromises data integrity, leading to suboptimal model performance or erroneous conclusions.

The heterogeneity of IoT devices further complicates matters by introducing variability in data preprocessing and model training. AI models may exhibit erratic behaviour without dependable data validation and normalisation mechanisms.

Security and Privacy Risks: As AI-powered IoT systems are often deployed in open, physically accessible, or resource-constrained environments, data security is a critical aspect. These systems are vulnerable to various cyber threats, such as data poisoning attacks (where malicious data is injected into the learning process) and unauthorised access to control or decision-making components.

Edge computing is widely used in IoT systems to reduce latency and increase throughput. However, they often lack the computing resources to support advanced security mechanisms.

Lack of Explainability and Transparency: A fundamental component of trust in AI is explainability. Unfortunately, many AI models used in IoT applications (e.g., deep learning networks) operate as "black boxes," making it challenging to trace decision paths or verify their reasoning. This opacity is problematic not only for end users but also for auditors, regulators, and system engineers. Continuous learning or online adaptation, often required in dynamic environments, further obscures model behaviour and complicates version control and auditing. Current regulatory frameworks increasingly mandate explainability and traceability, which IoT systems are not always equipped to provide.

Bias and Fairness: Uneven sensor placement, biased historical data, or algorithmic preferences in IoT data processing can lead to biased decisions, such as untimely maintenance or insufficient service coverage. In addition, sensor data may reflect historical inequalities or operational inconsistencies, and artificial intelligence models trained on such data may perpetuate or even reinforce these biases. Ensuring equity requires not only technical solutions but also ethical considerations in sensor placement, data collection policies, and algorithm development.

Reliability and Robustness: AI systems in IoT must operate reliably in real-world conditions, where noise, faults, or adversarial conditions are common. Models must be resilient to single-point failures (e.g., a sensor outage) and capable of operating with degraded inputs. However, many AI models are sensitive to small input perturbations and lack built-in fault tolerance mechanisms. Real-time requirements add another layer of complexity. AI models must make timely decisions, often with partial or delayed information. Any delay, crash, or misclassification in a mission-critical application (e.g., industrial automation or emergency health response) could lead to severe consequences. Ensuring robustness and real-time responsiveness is essential but technically challenging.

Legal and Ethical Ambiguities: The legal landscape for AI in IoT is still evolving. Questions around liability, accountability, and compliance are especially murky when autonomous systems act without direct human intervention. If a smart sensor triggers a harmful action based on a faulty AI decision, assigning legal responsibility can be complex. Ethical issues also abound. How should an AI prioritise actions in a resource-constrained situation? Who decides what constitutes a "fair" or "safe" outcome? In addition, the lack of universally accepted standards for trustworthy AI in IoT means developers often operate in regulatory grey areas, risking compliance violations and public distrust.

Addressing these trust challenges requires a holistic approach involving robust data governance, secure architectures, interpretable models, fairness audits, and alignment with evolving legal and ethical standards. The next section of this paper will explore proposed solutions and frameworks aimed at overcoming these barriers and enabling trustworthy AI in smart sensor systems.

The following conceptual model is proposed for the determination of the trustworthiness of artificial intelligence in the context of Internet of Things (IoT)-based smart sensor systems, encompassing a multitude of salient dimensions.

The Trustworthiness Score of an AI system in IoT is defined as follows:

$$T_{AI} = w_1 \cdot Q_D + w_2 \cdot R_A + w_3 \cdot S_C + w_4 \cdot P_V + w_5 \cdot E_X + w_6 \cdot F_B + w_7 \cdot C_L. \quad (1)$$

Where:

Q_D – Data Quality Score – Reflects completeness, accuracy, integrity, and timeliness of sensor data.

R_A – Robustness & Reliability – Measures system resilience under faults, noise, or adversarial inputs.

S_C – Security Confidence – Includes protection from cyberattacks, secure data transmission, etc.

P_V – Privacy Preservation Score – Reflects compliance with GDPR, CCPA, anonymisation practices, etc.

E_X – Explainability & Transparency – Measures clarity of AI decision-making and ability to audit.

F_B – Fairness & Bias Mitigation – Indicates bias resistance and equity in data/decisions.

C_L – Compliance & Legal Readiness – Alignment with legal standards, accountability, and documentation.

Weights can be domain-specific or adjusted based on risk profiles (e.g., safety-critical systems might assign higher w_2 or w_6)

$$w_i \in [0, 1], \text{ and } \sum w_i = 1.$$

Each factor Q_D, R_A, … should be evaluated on a **normalised scale** from 0 (low trust) to 1 (high trust), based on metrics, expert review, or benchmarks.

Example: For a manufacturing AI system:

- High data integrity: $Q_D = 0.9$
- Moderate explainability: $E_X = 0.6$
- Strong legal compliance: $C_L = 0.85$

With weights: $w_1 = 0.2$, $w_2 = 0.15$, $w_3 = 0.15$, $w_4 = 0.1$, $w_5 = 0.15$, $w_6 = 0.1$, $w_7 = 0.15$

Then, the overall T_{AI} can be computed according to Table 2.

Table 2

Computing of Total Trustworthiness Score

Factor	Symbol	Score (0-1)	Weight (0-1)	Weighted Score
Data Quality	Q_D	0,9	0,2	0,18
Robustness & Reliability	R_A	0,1	0,15	0,015
Security Confidence	S_C	0,4	0,15	0,06
Privacy Preservation	P_V	0,6	0,1	0,06
Explainability	E_X	0,6	0,15	0,09
Fairness & Bias	F_B	0,3	0,1	0,03
Legal Compliance	C_L	0,85	0,15	0,1275
Total Trustworthiness Score				**0,5625**

Solutions for Building Trustworthy AI in Smart Sensor Systems

In order to address the complex trust challenges present in AI-powered smart sensor systems, a combination of technical, organisational, and regulatory strategies is essential.

In the area of data governance and validation frameworks, it is advisable to use automated data quality checks in real-time. This will help to detect sensor drifts, anomalies and missing data in real-time. Sensor data logs should include calibration history, accuracy ratings, and environmental context. Redundant sensing will help improve data reliability.

To ensure data protection in Secure-by-Design Architectures, the implementation of end-to-end encryption is essential at both the edge sensor level and within the cloud environment. To optimise efficiency, it is crucial to employ resource-efficient security protocols, such as ECC and TLS-lite, particularly for devices with constrained capabilities. Limiting access through authentication at each system level is also important.

To ensure the explainability and transparency of AI models, it is essential to integrate explainable AI techniques such as SHAP, LIME, or decision trees to interpret the model's behaviour. Additionally, it is crucial to maintain audit trails of model inputs, outputs, and decisions over time. A human-in-the-loop (HITL) approach is recommended, wherein human operators can review critical AI decisions, such as anomaly detection and shutdown triggers.

It is advisable to use various data audit tools to ensure fairness and mitigate bias. The accuracy of the model's predictions should also be periodically assessed in terms of spatial, temporal, and demographic variables.

Increasing the robustness and reliability of models can be achieved by subjecting them to simulated disturbances or sensor failures to identify vulnerabilities, implementing self-healing mechanisms, and continuously training them.

To ensure legal, ethical, and regulatory compliance, it is essential to conduct structured assessments of how AI solutions impact people, the environment, and communities. Developments should align with key regulatory documents in the field, such as ISO/IEC 42001 (AI governance), the EU AI Act, and the NIST AI RMF.

Ensuring the clear display of information about certification, third-party audits, or regulatory compliance in the user interface will facilitate trusted communication and user engagement.

The following key solution pillars have been identified as instrumental in fostering trustworthy, resilient, and ethically aligned AI within such environments (Table 3).

Table 3

Solution mapping

Challenge	Solution Approach	Example Practice
Data Quality & Integrity	Redundant sensing, validation layers	Sensor fusion, data profiling tools
Security & Privacy	Secure-by-design principles	End-to-end encryption, zero trust architecture
Explainability	Use of XAI tools	SHAP/LIME visualisations, model cards
Fairness	Dataset audits & rebalancing	Bias detection tools, fairness metrics
Robustness	Adversarial testing, fallback logic	Fault simulation frameworks
Legal & Ethical	Compliance-by-design	Conformance to ISO/IEC 42001, EU AI Act

Summary and Conclusion

Intelligent sensor systems driven by artificial intelligence are becoming increasingly integrated into critical areas of human life, such as environmental monitoring, automated manufacturing, smart cities, and more. Given this integration, it is essential that the ability of these systems to make autonomous decisions is grounded in the principles of reliability, explainability, confidentiality, fairness, and legal responsibility. The research conducted in this article emphasizes the urgent need to establish trust in intelligent sensor systems controlled by artificial intelligence, especially in the context of the Internet of Things (IoT).

The development and implementation of key standards and regulations, such as the EU AI Act, the international standard ISO/IEC 42001, and the NIST AI RMF guidelines, underscores the significance of the issue and the multidimensional nature of trust in AI. Addressing concerns such as data quality, security risks, algorithmic bias, transparency, and regulatory uncertainty is crucial for ensuring the safe and ethical deployment of AI.

The proposed trustworthiness model and scoring framework offer a practical method for evaluating the trust level of AI systems in smart sensor environments. Furthermore, the identified solution pillars – ranging from explainable AI and secure architectures to fairness auditing and compliance-by-design – provide actionable guidance for developers, policymakers, and system integrators.

Achieving trust in artificial intelligence in the Internet of Things is not a one-time effort, but an ongoing interdisciplinary process that combines technical innovation with effective governance. Building trust at all levels – from data collection and model development to system oversight and user engagement – will ensure that AI technologies serve both industrial progress and social welfare.

Bibliography

1. Са́нчес П. М. С., Сельдра́н А. Х., Сє Н., Бове́ Ж., Пе́рес Г. М., Штї́ллер Б. (2024). Федеративна довіра: Рішення для надійного федеративного навчання. Future Generation Computer Systems, 152, 83–98. https://doi.org/10.1016/j.future.2023.10.013

2. Джагатісаперума́л С. К., Фам К'ю. Ві., Рубі Р., Ян Ч., Сюй Ц., Чжан Ч. (2022). Пояснюваний ШІ через Інтернет речей (IoT): Огляд, сучасний стан та майбутні напрямки. IEEE Open Journal of the Communications Society, 3, 2106–2136. https://doi.org/10.48550/arXiv.2211.01036

3. Насім, М. Д., Бісвас, П., Рашид, А., Бісвас, А., та Гупта, К. Д. (2024). Надійний XAI та його застосування. Препринт arXiv arXiv:2410.17139. https://doi.org/10.48550/arXiv.2410.17139

4. Лі, Б., Ці, П., Лю, Б., Ді, С., Лю, Дж., Пей, Дж., … та Чжоу, Б. (2023). Надійний ШІ: від принципів до практики. ACM Computing Surveys, 55(9), 1–46. https://doi.org/10.1145/3555803

5. Ван, Х., Ван, Б., Ву, Ю., Нін, З., Го, С., та Ю, Ф. Р. (2024). Опитування щодо надійного периферійного інтелекту: від безпеки та надійності до прозорості та сталого розвитку. IEEE Communications Surveys & Tutorials. https://doi.org/10.1109/COMST.2024.3446585

6. Раданлієв, П., Де Рур, Д., Мейпл, К., Нерс, Дж. Р., Ніколеску, Р., та Ані, У. (2024). Безпека штучного інтелекту та кіберризики в системах Інтернету речей. Frontiers in Big Data, 7, 1402745. https://doi.org/10.3389/fdata.2024.1402745

7. Черданцева, Ю., Бернап, П., Наджм-Техрані, С. та Джонс, К. (2022). Конфігурована модель залежностей SCADA-системи для цілеспрямованої оцінки ризиків. Appl. Sci. 12, 1–29. https://doi.org/10.3390/app12104880

8. Ерола, А., Аграфіотіс, І., Нерс, Дж. Р. К., Аксон, Л., Голдсміт, М. та Кріз, С. (2022). Система для розрахунку кіберцінності під ризиком. Comput. Secur. 113:102545. https://doi.org/10.1016/j.cose.2021.102545

9. Попеску, С. М., Мансур, С., Вані, О. А., Кумар, С. С., Шарма, В., Шарма, А., … & Чунг, Ю. С. (2024). Технології на основі штучного інтелекту та Інтернету речей для моніторингу та управління забрудненням навколишнього середовища. Frontiers in Environmental Science, 12. https://doi.org/10.3389/fenvs.2024.1336088

10. Ай, В., Чен, Г., Юе, Х., та Ван, Дж. (2023). Застосування гіперспектрального та глибокого навчання у виявленні мікропластику в ґрунті сільськогосподарських угідь. J. Hazard. Mater. 445, 130568. https://doi.org/10.1016/j.jhazmat.2022.130568

11. Аша, П., Натраян, Л., Гіта, Б., Беула, Дж. Р., Сумати, Р., Варалакшмі, Г. та ін. (2022). Інтернет речей, що дозволяє використовувати екологічну токсикологію для моніторингу забруднення повітря за допомогою методів штучного інтелекту. Environ. Res. 205, 112574. https://doi.org/0.1016/j.envres.2021.112574

References

1. Sánchez, P. M. S., Celdrán, A. H., Xie, N., Bovet, G., Pérez, G. M., & Stiller, B. (2024). Federated trust: A solution for trustworthy federated learning. *Future Generation Computer Systems*, 152, 83-98. https://doi.org/10.1016/j.future.2023.10.013

2. Jagatheesaperumal, S. K., Pham, Q. V., Ruby, R., Yang, Z., Xu, C., & Zhang, Z. (2022). Explainable AI over the Internet of Things (IoT): Overview, state-of-the-art and future directions. *IEEE Open Journal of the Communications Society*, 3, 2106–2136. https://doi.org/10.48550/arXiv.2211.01036

3. Nasim, M. D., Biswas, P., Rashid, A., Biswas, A., & Gupta, K. D. (2024). Trustworthy XAI and Application. *arXiv preprint arXiv*: 2410.17139. https://doi.org/10.48550/arXiv.2410.17139

4. Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J.,... & Zhou, B. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys*, 55(9), 1–46. https://doi.org/10.1145/3555803

5. Wang, X., Wang, B., Wu, Y., Ning, Z., Guo, S., & Yu, F. R. (2024). A survey on trustworthy edge intelligence: From security and reliability to transparency and sustainability. *IEEE Communications Surveys & Tutorials*. https://doi.org/10.1109/COMST.2024.3446585

6. Radanliev, P., De Roure, D., Maple, C., Nurse, J. R., Nicolescu, R., & Ani, U. (2024). AI security and cyber risk in IoT systems. *Frontiers in Big Data*, 7, 1402745. https://doi.org/10.3389/fdata.2024.1402745

7. Cherdantseva, Y., Burnap, P., Nadjm-Tehrani, S., and Jones, K. (2022). A configurable dependency model of a SCADA system for goal-oriented risk assessment. *Appl. Sci*. 12, 1–29. https://doi.org/10.3390/app12104880

8. Erola, A., Agrafiotis, I., Nurse, J. R. C., Axon, L., Goldsmith, M., and Creese, S. (2022). A system to calculate cyber value-at-risk. *Comput. Secur. 113*: 102545. https://doi.org/10.1016/j.cose.2021.102545

9. Popescu, S. M., Mansoor, S., Wani, O. A., Kumar, S. S., Sharma, V., Sharma, A., … & Chung, Y. S. (2024). Artificial intelligence and IoT driven technologies for environmental pollution monitoring and management. *Frontiers in Environmental Science*, 12. https://doi.org/10.3389/fenvs.2024.1336088

10. Ai, W., Chen, G., Yue, X., and Wang, J. (2023). Application of hyperspectral and deep learning in farmland soil microplastic detection. *J. Hazard. Mater*. 445, 130568. https://doi.org/10.1016/j.jhazmat.2022.130568

11. Asha, P., Natrayan, L., Geetha, B., Beulah, J. R., Sumathy, R., Varalakshmi, G., et al. (2022). IoT enabled environmental toxicology for air pollution monitoring using AI techniques. *Environ. Res*. 205, 112574. https://doi.org/0.1016/j.envres.2021.112574

Дата першого надходження рукопису до видання: 22.09.2025
Дата прийнятого до друку рукопису після рецензування: 16.10.2025
Дата публікації: 28.11.2025