VCDNP Vienna Center for Disarmament and Non-Proliferation

January 2026

Cyber Security in IoT/IIoT and AI Integration

Dr. Lobna Ben Khelifa

Vienna Center for Disarmament and Non-Proliferation | vcdnp.org

Author

Lobna Ben Khelifa is a PhD-qualified expert in Artificial Intelligence (AI) and the Internet of Things (IoT), with extensive international experience across academia, industry, and policy advisory roles in Europe, Asia, and the Middle East. She has led multiple projects in smart systems, secure IoT, and trustworthy AI. Currently, she serves as an Assistant Professor of Data Science and AI at Applied Science University in Amman, Jordan, and works as a consultant specialising in AI strategy, security, and emerging technologies.

Her work focuses on bridging research and practical implementation, with a strong emphasis on data security, trustworthy and generative AI, ethics and regulation, responsible innovation, security-driven AI and IoT systems, and the governance of emerging technologies.

About the VCDNP

The Vienna Center for Disarmament and Non-Proliferation (VCDNP) promotes international peace and security by conducting research, facilitating dialogue, and building capacity on nuclear non-proliferation and disarmament.

The VCDNP is an international non-governmental organisation, established in 2010 by the Federal Ministry for European and International Affairs of Austria and the James Martin Center for Nonproliferation Studies at the Middlebury Institute of International Studies at Monterey.

Our research and analysis provide policy recommendations for decision-makers. We host public events and facilitate constructive, results-oriented dialogue among governments, multilateral institutions, and civil society. Through in-person courses and online resources on nuclear non-proliferation and disarmament, we train diplomats and practitioners working in Vienna and around the world.

Acknowledgements

This research and paper were made possible through the support of the Vienna Center for Disarmament and Non-Proliferation (VCDNP) as well as a research project funded by **Global Affairs Canada**.

VCDNP
Vienna Center for Disarmament and Non-Proliferation

Andromeda Tower, 13/1
Donau-City-Strasse 6
1220 Vienna
Austria

vcdnp.org
info@vcdnp.org
@VCDNP
VCDNP

Sponsored by
Canada

Contents

Recent innovations in AI have made IoT ecosystems more resilient, adaptive, and capable of predictive maintenance, personalised services, and decision-making support.

Introduction

The concept of the Internet of Things (IoT) has evolved remarkably since its origins. First introduced by Kevin Ashton in 1999, it began as a vision for supply chain logistics linking radio-frequency identification technology to the Internet so that physical objects could be tracked and managed automatically, without human intervention.[1] Over time, this narrow idea expanded into a global vision, formalised by organisations such as the International Telecommunication Union and Internet Engineering Task Force, of a networked infrastructure of smart, connected objects that may or may not be connected to the Internet.[2]

A useful distinction can be made between consumer IoT and the Industrial Internet of Things (IIoT). Consumer IoT focuses on enhancing everyday life, powering applications such as smart homes and wearable technology that brings greater convenience and automation. IIoT, by contrast, is tailored for industrial sectors such as manufacturing, energy, and logistics. It integrates operational and information technologies to optimise complex processes, improve safety, and maximise efficiency. Compared to consumer IoT, IIoT demands higher reliability, precision, and stronger security, often relying on specialised communication protocols. The interconnected nature of IIoT exposes industrial systems to new vulnerabilities, such as heightened cyber security risks, that can disrupt operations or compromise safety, making robust protection a top priority.[3]

1 Kevin Ashton, "That 'Internet of Things' Thing", RFiD Journal, 22 Jun 2009. Available at: https://www.rfidjournal.com/expert-views/that-internet-of-things-thing/73881/.

2 International Telecommunication Union , "Overview of the Internet of things", Series Y: Global Information Infrastructure, Internet Protocol Aspects and Next-Generation Networks, Y.2060, Jun 2012. Available at: https://handle.itu.int/11.1002/1000/11559.

3 Pai Zheng, Xun Xu, and Lihui Wang, "Industrial Internet-of-Things (IIoT)-enabled digital servitisation", International Journal of Production Research, Vol. 61, Issue 12, 27 May 2023. Available at: https://www.tandfonline.com/doi/full/10.1080/00207543.2023.2202258.

What was once about merely connecting devices has now matured into a complex system that encompasses edge computing,[4] artificial intelligence (AI), and combined cyber-physical systems. The most transformative stage of this evolution has been the convergence of IoT with AI and machine learning (ML), a fusion often referred to as the Artificial Intelligence of Things (AIoT) or the Internet of Smart Things. Unlike traditional IoT, which mainly collects and transmits data, AIoT systems can analyse, learn, and even act autonomously. For instance, a conventional IoT camera may only stream video, whereas an AIoT camera can recognise a specific individual, detect anomalies, and trigger an immediate response. This shift elevates IoT from simple connectivity to intelligent, decision-making ecosystems.

However, the growing complexity of IoT ecosystems also deepens their cyber security challenges.[5] Millions of devices now generate massive volumes of data across distributed networks, creating opportunities for intrusion and attack. In addition to enabling AIoT, AI models and systems play an increasingly critical role in securing and managing IoT networks. By analysing vast streams of real-time data, AI models and systems can detect anomalies, predict threats, and automatically respond to security incidents with minimal human intervention. Machine learning and deep learning enhance this further by identifying previously unknown cyber threats and unusual behaviours. Beyond security, AI models and systems strengthen access control through behavioural authentication, support privacy-preserving approaches such as federated learning,[6] and boost efficiency through edge computing. Collectively, these innovations make IoT ecosystems more resilient, adaptive, and capable of predictive maintenance, personalised services, and decision-making support.

4 Edge computing processes data locally, at or near the IoT device. A full explanation of the term appears on page 5.

5 Assad Abbas, "Artificial Intelligence-Driven Cyber Security in IoT: Ensuring Secure Connections in an Interconnected World", Atlantic.net, 22 May 2025. Available at: https://www.atlantic.net/gpu-server-hosting/artificial-intelligence-driven-cyber security-in-iot-ensuring-secure-connections-in-an-interconnected-world/.

6 Federated learning is a machine learning technique that trains AI models locally on devices without sharing raw data with a central server. This is discussed in more detail on page 20.

IoT allows for real-time monitoring and automation, enabling efficiency, safety, and responsiveness.

Understanding the Internet of Things (IoT) and the Industrial Internet of Things (IIoT)

Key Components of IoT Systems

IoT Layered Architecture

The architecture of IoT can be understood through four key layers, with an additional cross-cutting security layer (illustrated in Figure 1 below).

The **Sensing Layer**, also known as the perception layer, is responsible for capturing data from the physical environment through sensors, actuators, radio-frequency identification tags, and other embedded devices. It acts as the sensory system of IoT/IIoT, converting real-world parameters such as temperature, motion, and location into digital signals.

The **Processing Layer** manages, stores, and analyses the data collected by the sensing layer. Often referred to as the middleware or data layer, it incorporates cloud and edge computing, big data analytics, and machine learning to generate insights and support decision-making. This layer also handles governance, operations, and management of the overall IoT/IIoT system, ensuring compliance with policies, efficient workflows, and system reliability through monitoring, updates, and resource management.[7]

7 Ilya Katlinsky, "IoT architecture: key layers, components & use cases", itransition, 10 Apr 2025. Available at: https://www.itransition.com/iot/architecture.

The **Communication Layer**, also called the network layer, enables the secure and reliable transfer of data between devices, processing systems, and applications. It relies on a variety of communication networks, protocols, gateways, and internet technologies to ensure seamless connectivity and interoperability across diverse IoT/IIoT environments.

The **Application Layer** delivers services to end users through intuitive interfaces, data visualisation tools, and dashboards. It provides the means for device control, insight presentation, and integration with enterprise systems via application programming interfaces, enabling IoT/IIoT technologies to support business operations and decision-making processes.

Finally, the **Security Layer** functions as a cross-cutting component that protects every stage of the architecture. It ensures the confidentiality, integrity, and availability of IoT/IIoT systems by applying appropriate security measures at each level, from device protection to secure communication and controlled access to applications.

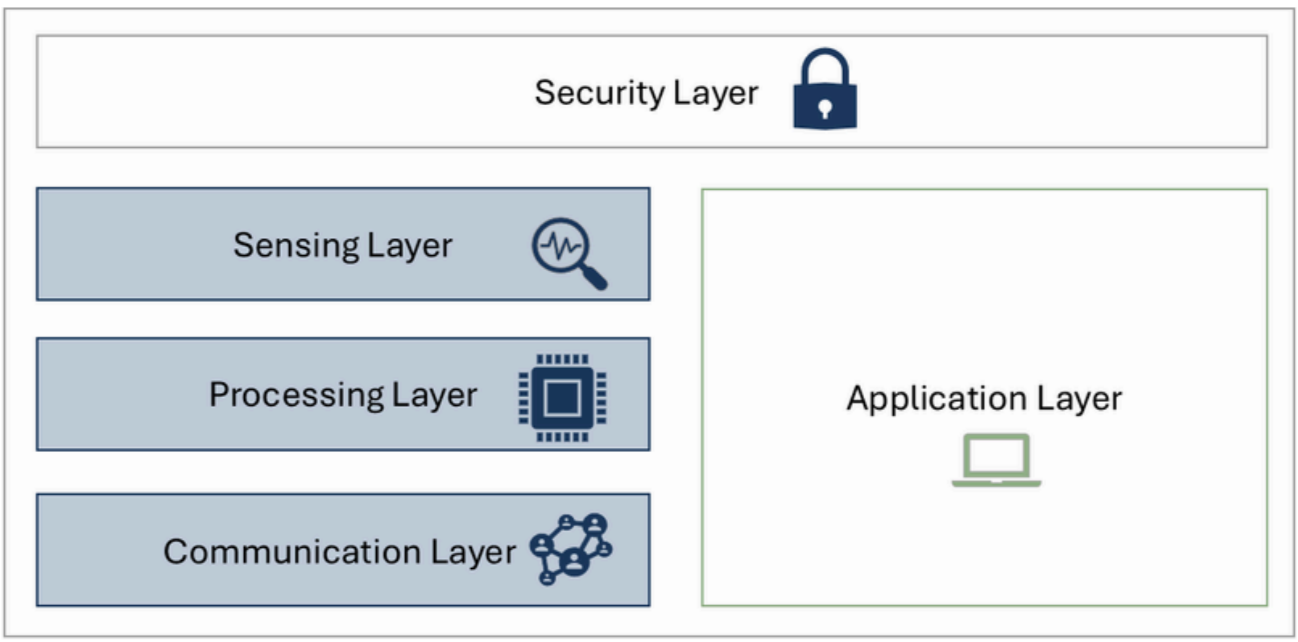

Figure 1. IoT Layers

IoT Functional Components

Within this architecture, four components are essential:

1. **Data Component:** The raw information collected by sensors and devices at the sensing layer.
2. **Analytics Component:** The machine learning algorithms to analyse data, extract insights, and automate decision-making primarily at the processing layer.
3. **Connectivity Component:** The networks and protocols (such as Wi-Fi, Bluetooth, 5G) that enable data transfer between devices and systems at the communication layer.
4. **Physical Resources:** The hardware, such as sensors, actuators, and embedded devices, scattered across the sensing and communication layers, enabling interaction with the physical environment.

Table 1 outlines the IoT layers, their core functions, and the corresponding functional components that support each layer.

Table 1. IoT Layers and Linked Functional Components

IoT Layer	Main Function	Linked Functional Component(s)
Sensing (Perception)	Capture real-world data	Physical Resources + Data Component
Communication (Network)	Enable reliable bidirectional data transfer	Connectivity Component (supports both upward data flow and downward commands)
Processing (Data/Middleware)	Store, process, and analyse data	Data Component + Analytics Component
Security (Cross-cutting)	Protect devices, data, and operations	All Components
Application	Present insights and control devices	Analytics Component (visualisation, decisions) + Communication (bidirectional commands/data)

Data Sharing and Remote Control in IoT

Real-Time Monitoring and Automation

At its core, IoT is about using raw sensor data and turning it into action. Real-time monitoring, which involves the continuous and immediate collection of data from connected devices, provides a live, unfiltered view of a system's status, whether it is temperature, the pressure inside a pipe, or the operational status of a factory robot.[8] This constant stream of information is the fuel for automation. Once an AIoT system detects a specific condition, like a sudden drop in pressure or a temperature spike, it can automatically trigger a pre-programmed response without human intervention. For instance, an automated system in a smart factory could respond to an overheating machine by shutting it down to prevent a catastrophic failure. A seamless loop of sensing, analysing, and acting is the key factor contributing to the effectiveness of IoT systems, enabling unprecedented efficiency, safety, and responsiveness across countless industries.

Cloud vs. Edge Computing in IoT

The effectiveness of real-time monitoring and automation depends heavily on where the data is processed. This can be done either remotely or locally, via cloud computing or edge computing. **Cloud computing** relies on a centralised data centre, located remotely, to handle all processing and storage. Devices send raw data to the cloud, where powerful servers perform complex analytics and store vast amounts of information.[9]

However, with advancements in cloud infrastructure and high-speed connectivity, cloud computing is increasingly capable of supporting real-time processing as well, enabling applications such as online gaming, video streaming, and real-time data analytics. The scalability and flexibility of the cloud make it possible to dynamically allocate computing resources as needed, ensuring responsiveness even for time-sensitive tasks.

One of the main challenges of this model remains latency, caused by the time it takes for data to travel back and forth between devices and the cloud. To address this, **edge computing** processes data locally, at or near the device, allowing systems to make rapid, autonomous decisions within milliseconds – essential for applications such as self-driving cars, industrial automation, or smart medical devices.[10]

8 World Economic Forum, “Industrial Internet of Things: Unleashing the Potential of Connected Products and Services”, Jan 2015. Available at: https://www3.weforum.org/docs/WEFUSA_IndustrialInternet_Report2015.pdf.

9 Justyna, "Edge Computing vs Cloud Computing: Deciding How to Handle Your IoT Data in the Enterprise," Multishoring, 28 Apr 2025. Available at: https://multishoring.com/blog/edge-computing-vs-cloud-computing/.

10 David, “Edge Computing vs. Cloud Computing. Which option is better for IoT projects?“ DeepSea Developments, Blog. Available at: https://www.deepseadev.com/en/blog/edge-computing-vs-cloud-computing/.

Edge computing also reduces bandwidth usage by sending only filtered or critical data to the cloud for long-term storage or further analysis. In many cases, a **hybrid approach** offers the best solution, combining the real-time responsiveness of edge computing with the scalability and massive storage capacity of the cloud.[11]

Moreover, storing data in the cloud raises important concerns about data security and sovereignty. Sensitive information may be stored in data centres located in different countries, raising questions about who controls access and how the data is protected from breaches or misuse. Therefore, organisations must enforce strict measures for encryption, identity and access management, and ensure compliance with national and international data protection regulations.

Security Implications of Remote Access

While remote access provides incredible convenience, it also creates significant security risks that need careful management. Every connection point is a potential entry for a cyberattack, and a single vulnerability can compromise an entire system. A major problem is weak authentication, as many IoT devices come with default or easily guessable passwords that are rarely changed. A lack of robust access control makes it simple for an attacker to gain unauthorised entry. Furthermore, data interception is a critical concern. Data transmitted between a device and a remote user is vulnerable to a "man-in-the-middle" attack if it is not properly encrypted. An attacker could not only steal sensitive information but also inject malicious commands to hijack the device. Another persistent issue is the lack of regular updates. Many IoT devices are not designed to receive regular security patches, leaving them permanently vulnerable to new exploits as they emerge. To protect an IoT ecosystem, it is crucial to implement strong passwords, multi-factor authentication, end-to-end data encryption, and regular security audits.

Scalability Challenges in IoT Networks

The fundamental promise of IoT is to connect billions, even trillions, of devices. However, this very promise presents its most significant hurdle: scalability. A network is considered scalable if it can maintain its performance as more resources are added to accommodate increased workloads. Traditional internet architectures struggle under the unique demands of massive IoT deployments, leading to two core challenges: managing large-scale deployments and overcoming interoperability issues.

Managing Large-Scale Deployments

This challenge concerns the technical and logistical difficulties in expanding an IoT system to accommodate a rapidly increasing number of devices without degrading performance, security, or manageability.

The core challenge of managing large-scale deployments is the immense strain that millions of devices place on network bandwidth, data infrastructure, and management systems. It requires a fundamental shift from traditional IT models to automated, efficient, and lightweight architectures to avoid crippling congestion, exorbitant costs, and unmanageable complexity. Table 2 below outlines the key areas of impact and provides concrete examples of the challenges faced during large-scale IoT deployment.

11 NTSO-E, “Cloud and Edge Computing“, ENSTO-E Technopedia, 25 Mar 2025. Available at: https://www.entsoe.eu/technopedia/techsheets/cloud-and-edge-computing/.

Table 2. Examples of Challenges in Managing Large-Scale Deployments

Challenge	Problem	Solution
Network Capacity and Congestion	IoT devices generate massive, often simultaneous, bursts of data (e.g. millions of smart meters reporting energy usage at the top of the hour). This can overwhelm traditional cellular (4G/5G) and Wi-Fi networks, causing congestion, packet loss, and increased latency.	Lightweight protocols like **Message Queuing Telemetry Transport** and **Constrained Application Protocol** are designed specifically for this purpose. They use a publish-subscribe model and are header-efficient, drastically reducing network overhead.[12]
Data Management and Storage	The volume, velocity, and variety of data generated by a scalable IoT network are immense. Storing every single data point in a raw format in a central cloud database is prohibitively expensive and inefficient for querying and analysis.	**Edge Computing**: Process and filter data locally on gateways or devices before sending only relevant insights or aggregated data to the cloud. This reduces bandwidth and storage costs.[13] **Cloud Data Platforms**: Use scalable cloud data lakes and warehouses (e.g. Amazon Web Services IoT Analytics, Azure Time Series Insights) designed to handle time-series IoT data efficiently.
Device Management and Provisioning	Manually configuring and securing each device in a deployment of millions is impossible. Tasks like software updates (over-the-air updates), security credential rotation, and device monitoring must be fully automated.	Robust IoT device management platforms (e.g. Amazon Web Services IoT Core Device Management, Azure IoT Hub) are essential. They provide automated registration (onboarding), group-based policy management, and remote monitoring to manage devices at scale.[14]
Security and Identity Management	Each device is a potential entry point for an attack. Managing unique cryptographic keys and certificates for billions of devices, and ensuring they can be rotated and revoked, is a monumental task.	Automated certificate provisioning (e.g. using a trust-on-first-use model) and lightweight cryptographic standards are critical for scalable security.

Interoperability Issues

Interoperability is the ability of different systems, devices, and applications from various vendors to work together seamlessly by exchanging and using data. Its absence is a primary barrier to achieving large-scale, heterogeneous IoT ecosystems.

The challenge of interoperability stems from a fragmented technological landscape where devices speak different "languages" (protocols) and use different "dialects" (data models). Without common standards, IoT systems risk becoming isolated silos, preventing seamless data exchange and integrated functionality that unlocks the full value of a connected world. Table 3 below presents examples of these challenges and provides possible solutions.

12 Andy Stanford-Clark and Hong Linh Truong, "MQTT For Sensor Networks (MQTT-SN): Protocol Specification", Version 1.2, International Business Machines Corporation (IBM), 14 Nov 2013.

13 Weisong Shi, et al., "Edge Computing: Vision and Challenges", IEEE Internet of Things Journal, Vol. 3, Issue 5, Oct. 2016. Available at: https://doi.org/10.1109/JIOT.2016.2579198.

14 Amazon Web Service (AWS), "AWS IoT Device Management". Available at: https://aws.amazon.com/iot-device-management/.

Table 3. Examples of Challenges of Interoperability

Challenge	Problem	Solution
Protocol Heterogeneity	The IoT landscape is fragmented with countless communication protocols. Devices may use Bluetooth for short-range, Long Range Wide Area Network for long-range, low-power, Zigbee for mesh networks, and message queuing telemetry transport for cloud communication. Getting these different protocols to work together is complex.	**IoT Gateways** are a primary solution. They act as translators, connecting to devices using various local protocols, processing the data, and then forwarding it to the cloud using a standard language like message queuing telemetry transport or hypertext transfer protocol.[15]
Data Semantics and Format	Even if two devices can connect, they might speak different "data languages". For example, one sensor might send temperature data as {"temp": 23} while another uses {"temperature_celsius": 23}. Without a common schema, applications cannot universally understand the data.	Adopting **standardised data models** is key. Initiatives like **Semantic Web Technologies JavaScript Object Notation for Linked Data,** the **World Wide Web Consortium Web of Things Thing Description**, and industry-specific standards (e.g. Open Platform Communications - Unified Architecture for industrial automation) provide common vocabularies and frameworks to ensure data is interpreted correctly across systems.[16]
Platform Silos	Major IoT cloud providers (Amazon Web Services, Azure, Google Cloud) have their own ecosystems. Devices and applications designed for one platform often do not integrate easily with another, locking users into a single vendor and hindering the creation of broader solutions.	The movement towards **open standards** and **open-source frameworks** is combating this. Efforts like the **Eclipse Foundation** projects (e.g. Eclipse Ditto, Eclipse Hono) provide vendor-neutral middleware for building interoperable IoT solutions.[17]

15 Ala Al-Fuqaha, et al., “Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications”, IEEE Communications Surveys & Tutorials, Vol. 17, Issue 4, 15 Jun 2015, pp. 2347 - 2376. Available at: https://doi.org/10.1109/COMST.2015.2444095.

16 World Wide Web Consortium, “Web of Things (WoT) Thing Description,” W3C Recommendation, 9 Apr 2020. Available at: https://www.w3.org/TR/2020/REC-wot-thing-description-20200409/.

17 For more details on efforts by the Eclipse Foundation see its website at: https://iot.eclipse.org/.

The convergence of the digital and physical worlds means that cyberattacks can now have direct, and sometimes catastrophic, real-world consequences.

Cyber Security Challenges in IoT/IIoT

The exponential growth of IoT and IIoT has created a vast and complex attack surface. Security in these systems is indeed a chain, and its strength is determined by its weakest link. The convergence of the digital and physical worlds means that cyberattacks can now have direct, and sometimes catastrophic, real-world consequences.

Common Vulnerabilities in IoT/IIoT Systems

IoT and IIoT systems share several common vulnerabilities arising from their interconnected and resource-constrained nature. Weak authentication mechanisms, limited security features in low-cost devices, a lack of standardised protocols, and insufficient patch management often leave these systems exposed. Additionally, their reliance on continuous connectivity expands the potential attack surface, while the integration of legacy technologies in industrial settings further increases risks. These vulnerabilities collectively make IoT and IIoT environments attractive targets and highlight the need for stronger, systemic security measures.

Types of Attacks in IoT/IIoT Systems

Attacks on IoT and IIoT devices can be analysed across four fundamental components: data, models, communication, and physical resources. At the data level, adversaries may attempt to undermine its confidentiality, integrity, or availability. In models, threats target the processes that enable intelligent decision-making, seeking to distort or degrade their reliability. Communication layers are exposed to risks that aim to compromise the secure and consistent exchange of information between interconnected devices.

At the physical resources level, attackers target the hardware and infrastructure, threatening the stability and operational continuity of the entire system. Together, these dimensions highlight the broad and complex attack surface inherent in IoT and IIoT environments. Table 4 outlines a non-exhaustive taxonomy of attacks that could target different components of the system.

Table 4. Example of Attacks That Can Target Every Component of the IoT/IIoT Architecture Stack

Attacks on Data	**Attacks on Analytics Component (AI/ML)**
1. Data Integrity Attacks: Altering data in transit or at rest. For example, an attacker could manipulate sensor data from a temperature monitor to force a system into an unsafe state or falsify financial meter readings. 2. Data Theft: Exfiltrating sensitive collected data (e.g. personal health information from a wearable device or proprietary production formulas from an IIoT sensor). 3. Data Poisoning: Compromising the training data causing the AI model to learn incorrect patterns and behave maliciously once deployed.	1. Adversarial Machine Learning: Feeding manipulated input data to an IoT system's AI model to cause it to make incorrect decisions. For example, subtly altering an image to cause a computer vision system in a security camera to misclassify an intruder as a benign object. 2. Model Inversion: Attempting to reverse-engineer the training data from a deployed ML model, potentially reconstructing sensitive information about individuals or operations.
Attacks on the Connectivity Component	**Attacks on Physical Resources**
1. Denial-of-Service / Distributed Denial-of-Service: Overwhelming devices or network bandwidth with traffic, rendering them unresponsive. The Mirai botnet famously hijacked thousands of weak IoT devices to launch massive distributed denial-of-service attacks.[18] 2. Man-in-the-Middle: Intercepting and potentially altering communication between two parties without their knowledge (e.g. between a sensor and a controller).	1. Physical Tampering: Gaining physical access to a device to extract data, firmware, or cryptographic keys; or to modify its hardware. 2. Resource Depletion: Draining a device's battery life through constant communication requests, rendering it useless (a type of denial-of-service attack).

Defence Techniques

Securing IoT and IIoT requires a parallel focus on the same four components: data, models, communication, and physical resources. At the data level, measures are needed to preserve trustworthiness and resilience against manipulation. For models, the emphasis is on maintaining accuracy, transparency, and reliability in their operation despite potential adversarial pressures.

In communication channels, security measures must ensure that connectivity remains both protected and dependable. At the physical layer, protections are required to uphold the robustness and availability of the underlying devices and infrastructure. Addressing all four areas in a unified manner provides a comprehensive framework for strengthening the resilience and sustainability of IoT and IIoT ecosystems.

A defence-in-depth strategy is crucial, involving multiple layers of security measures that provide redundancy and increase capacity to delay, detect, and respond to an attack. Table 5 outlines potential mitigation strategies to defend the various system components.

18 More details about the Mirai Botnet can be found on radware, “What is the Mirai Botnet?” at https://www.radware.com/security/ddos-knowledge-center/ddospedia/mirai/.

Table 5. Mitigation Strategies and Defence Mechanisms for Securing IoT/IIoT System Components

Data Security • **Encryption at rest and in transit** (advanced encryption standard, transport layer security) to prevent unauthorised access. • **Access control and authentication** to ensure only authorised users/apps can view or modify data. • **Data anonymisation and masking** for sensitive personal information. • **Regular backups and secure storage** to mitigate loss or corruption. • **Integrity checks** (hashing,[19] digital signatures) to detect tampering.	**Model Security (AI/ML Models in IoT/IIoT)** • **Adversarial training** to improve robustness against malicious inputs. • **Model encryption and watermarking** to protect intellectual property. • **Access restriction** to prevent model theft or unauthorised retraining. • **Continuous monitoring of outputs** to detect abnormal behaviour or poisoning attacks. • **Regular updates of models** to address vulnerabilities and adapt to new threats.
Communication Security • **End-to-end encryption** ((datagram) transport layer security, virtual private networks) for data exchanged between devices and servers. • **Secure routing protocols** to prevent spoofing, replay, or man-in-the-middle attacks. • **Mutual authentication** between devices and gateways. • **Network segmentation** to isolate IoT/IIoT devices from critical systems. • **Intrusion detection and anomaly monitoring** to identify suspicious traffic patterns.	**Physical Resources Security (Devices & Infrastructure)** • **Tamper-resistant hardware design** (secure chips, trusted execution environments). • **Physical access control** (locks, surveillance, restricted zones). • **Firmware integrity checks** to prevent malicious modifications. • **Regular patching and updates** to address vulnerabilities in device software. • **Energy/resource management strategies** to protect against resource exhaustion (e.g. battery-draining attacks, denial-of-service).

Threats Specific to Industrial IoT (IIoT)

IIoT inherits all general IoT risks but faces heightened stakes due to its direct connection to the physical world.

Risks to Critical Infrastructure (OT/IT Convergence)

The historical separation between Information Technology (IT) systems (for data processing) and Operational Technology (OT) systems (for controlling industrial processes) is dissolving. This convergence creates new risks:

- **Expanded Attack Surface:** An infection on a corporate IT network (e.g. via a phishing email) may serve as a pivot point to access and attack critical OT networks, such as controlling machinery, power grids, or water treatment plants.
- **Incompatible Security Postures:** Many industrial OT systems are often old, fragile, and cannot be patched frequently (if at all). They were designed for reliability and safety, not security, and are often incompatible with standard IT security tools. A simple scan can crash a critical programmable logic controller.
- **Catastrophic Consequences:** Attacks can lead to physical damage, environmental disasters (e.g. chemical spills), prolonged production downtime costing millions, and even threats to human life.

19 "Hashing" refers to a one-way mathematical function that creates a unique "fingerprint" or "hash" for data that cannot be reversed or decoded. Any alteration would cause a new data "fingerprint" or "hash" to be generated that would not match the original.

Case Study "Stuxnet (2010)": A sophisticated worm specifically designed to target Siemens supervisory control and data acquisition systems and damage Iran's uranium enrichment infrastructure. It demonstrated the potential for a cyber weapon to cause physical destruction by reprogramming programmable logic controllers to spin centrifuges out of control while hiding the changes from operators.[20]

Supply Chain Attacks

The complex, global nature of IIoT supply chains introduces critical vulnerabilities.

- **Compromised Hardware/Software**: A malicious component (e.g. a backdoored chip, a vulnerable software library) introduced by a supplier can infect an entire product line or industrial system.
- **Third-Party Service Providers**: Attackers can target less-secure vendors (e.g. a maintenance firm with remote access to industrial systems) as a stepping stone to their ultimate target.
- **Lack of Visibility**: Organisations often have limited visibility into their suppliers' security practices, making it difficult to assess and mitigate these risks.

Case Study "The SolarWinds Attack (2020)": While not exclusively an IIoT incident, it is a quintessential supply chain attack. Nation-state actors compromised the software update mechanism of SolarWinds' Orion IT management platform. This malicious update was then distributed to ~18,000 customers, including critical government agencies and infrastructure operators, granting the attackers a foothold in their networks.[21]

20 Ralph Langner, "Stuxnet: Dissecting a Cyberwarfare Weapon", IEEE Security and Privacy, Vol. 9, Issue 3, 23 May 2011, pp. 49 -51. Available at: https://ieeexplore.ieee.org/document/5772960.

21 IBM, "SolarWinds Orion (CVE-2020-10148)", IBM Security Randori, last updated 10 May 2024. Available at: https://www.ibm.com/docs/en/randori?topic=2022-solarwinds-orion-cve-2020-10148.

AI systems can help detect anomalies associated with intrusions, malware propagation, denial-of-service attempts, or lateral movement within IoT networks.

The Role of AI in Enhancing IoT/IIoT Security

The scale, complexity, and unique vulnerabilities of IoT/IIoT ecosystems make traditional, human-centric security approaches insufficient. AI models and ML techniques are becoming critical force multipliers, enabling a shift from reactive to **proactive and adaptive** cyber security defence.

AI-Driven Threat Detection and Anomaly Detection

Traditional signature-based detection methods are often insufficient against novel or sophisticated attacks targeting IoT devices.[22,23] AI systems offer a more robust approach by learning normal device and network behaviour and identifying subtle deviations that may indicate malicious activity.[24,25] By continuously analysing telemetry data and network traffic, AI systems can detect anomalies associated with intrusions, malware propagation, denial-of-service attempts, or lateral movement within IoT networks.[26,27]

22 European Union Agency for Cyber Security (ENISA), "ENISA Threat Landscape 2023", ENISA, 19 Oct 2023. Available at: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023.

23 Michael Fagan et al., "IoT Device Cyber Security Guidance for the Federal Government: IoT Device Cyber Security Requirement Catalog", NIST Special Publication 800-213A, U.S. Department of Commerce, Nov 2021. Available at: https://doi.org/10.6028/NIST.SP.800-213A.

24 Tanish Baranwal et al., "Machine Learning-Based Anomaly Detection of Correlated Sensor Data: An Integrated Principal Component Analysis-Autoencoder Approach", arXiv preprint arXi2505.24044, 29 May 2025. Available at: https://arxiv.org/html/2505.24044v1.

25 Emanuel Krzysztoń, Izabela Rojeck, and Dariusz Mikolajewski, "Comparative Analysis of Anomaly Detection Methods in IoT Networks: An Experimental Study", Applied Sciences, Vol. 14, No. 24, 11 Dec 2024. Available at: https://www.mdpi.com/2076-3417/14/24/11545.

26 Ibid.

27 United Nations Peacekeeping, "Strategy for the Digital Transformation of UN Peacekeeping", United Nations, Sep 2021. Available at: https://peacekeeping.un.org/en/strategy-digital-transformation-of-un-peacekeeping.

Recent studies have demonstrated that AI-driven anomaly detection achieves higher accuracy and lower false-positive rates compared to conventional methods, making it especially effective in dynamic and noisy IoT environments.[28] AI-driven systems are able to adapt to changing operational patterns and scale across large fleets of heterogeneous devices, which is crucial in modern IoT ecosystems. While anomaly detection identifies deviations, ML extends this by modelling behaviour over time.

Machine Learning for Behavioural Analysis

Machine learning plays a pivotal role in enhancing IoT security by continuously modelling and monitoring device and network behaviour. By analysing large volumes of telemetry, such as central processing unit usage, memory utilisation, communication frequency, packet sizes, and destination addresses, ML establishes unique behavioural baselines for each device type and the network as a whole.[29] Once this baseline is defined, the system can identify abnormal activity in real time, which may help in identifying potential attacks, including those that do not match known signatures.

Behavioural analysis using ML enables the detection of a wide range of threats, including zero-day attacks, insider threats, compromised devices, and subtle deviations caused by malware or misconfiguration.[30] For example, ML could help notify of unusual behaviour such as:

- A temperature sensor that suddenly attempts to communicate with an unknown external server.
- A video surveillance camera transmitting data at unusual times when it should be inactive.
- A programmable logic controller in an industrial control system is issuing commands at a frequency or sequence inconsistent with its normal operation.

Beyond anomaly detection, ML-based behavioural analysis supports risk scoring and prioritisation by quantifying how far a device's behaviour deviates from its normal baseline. This allows security teams to focus on the most suspicious devices and network segments, improving operational efficiency.[31]

Moreover, behavioural ML models can adapt to dynamic IoT environments. Techniques such as unsupervised learning, clustering, and ensemble methods allow the models to handle concept drift, heterogeneous devices, and noisy telemetry data without extensive human supervision. These adaptive capabilities are particularly valuable for large-scale IoT deployments, such as smart cities, industrial automation, and healthcare IoT, where devices continuously generate massive amounts of diverse data.

Studies have consistently shown that ML-driven behavioural analysis reduces false positives, improves detection accuracy, and uncovers malicious activity that traditional signature-based systems would miss.[32,33] Thus, integrating ML-driven behavioural monitoring into IoT security strategies would assist organisations in achieving real-time threat detection while maintaining operational continuity.

28 Emanuel Krzysztoń et al., "Comparative Analysis of Anomaly Detection Methods in IoT Networks: An Experimental Study", Applied Sciences, Vol. 14, No. 24, 11 Dec 2024. Available at: https://www.mdpi.com/2076-3417/14/24/11545.

29 United Nations Peacekeeping, "Strategy for the Digital Transformation of UN Peacekeeping", United Nations, Sep 2021. Available at: https://peacekeeping.un.org/en/strategy-digital-transformation-of-un-peacekeeping.

30 United Nations, "E-Government Survey 2024: Accelerating Digital Transformation for Sustainable Development With the addendum on Artificial Intelligence", Department of Economic and Social Affairs, 2024. Available at: https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2024.

31 United Nations, "Resource Guide on Artificial Intelligence Strategies", United Nations, Apr 2021. Available at: https://sdgs.un.org/sites/default/files/2021-04/Resource%20Guide%20on%20AI%20Strategies_April%202021_rev.pdf.

32 Emanuel Krzysztoń et al., "Comparative Analysis of Anomaly Detection Methods in IoT Networks: An Experimental Study", Applied Sciences, Vol. 14, No. 24, 11 Dec 2024. Available at: https://www.mdpi.com/2076-3417/14/24/11545.

33 United Nations Peacekeeping, "Strategy for the Digital Transformation of UN Peacekeeping", Sep 2021. Available at: https://peacekeeping.un.org/en/strategy-digital-transformation-of-un-peacekeeping.

Predictive Security with AI Models

Beyond detecting active threats, AI systems play a crucial role in forecasting emerging attack surfaces and vulnerabilities within IoT networks. By analysing device and network telemetry alongside threat intelligence, AI systems can predict which nodes or links are most likely to be targeted, allowing security teams to prioritise where to focus preventive measures such as patching, access restrictions, and configuration hardening before exploitation occurs.[34]

Recent reports (2023–2024) highlight the rise of attacker "as-a-service" ecosystems[35] and evolving techniques, emphasising the need for predictive analytics that integrate threat intelligence with local IoT data.[36] This approach enables anticipation of attacks such as distributed denial-of-service waves, ransomware staging, and supply-chain compromises providing proactive defence rather than reactive response.

AI-driven predictive security involves several key strategies:

1. **Vulnerability Prediction:** By analysing device firmware, configuration files, and network topologies, AI systems can assess which assets are most likely to be exploited, taking into account known weaknesses, historical attack trends, and operational criticality.

2. **Threat Forecasting:** Natural language processing techniques allow AI models and systems to monitor hacker forums, dark web marketplaces, and open-source news feeds for early mentions of IoT exploits or planned attacks, providing actionable early warning for security teams.[37]

Enabling proactive measures with AI assistance allows organisations to fortify systems before attacks occur and reduce the potential impact of cyber threats. This predictive capability is particularly valuable in large-scale IoT deployments, where traditional reactive approaches may fail to keep pace with rapidly evolving threats.

AI-Powered Automated Responses

When coupled to policy engines[38] and security orchestration, automation, and response like playbooks,[39] AI systems can drive closed-loop actions – quarantining suspicious devices, throttling anomalous traffic, rotating credentials/keys, or enforcing dynamic micro-segmentation – while reducing the amount of time that systems, devices, or networks are unavailable or disrupted due to a security incident. For example, if an AI system detects suspicious activity on a device, it can quarantine the device or throttle traffic without shutting down the entire network.

Research prototypes pair anomaly detection with automated mitigations, and sector studies show that several industries are actively exploring this approach. For example, the energy sector, including smart grids and renewable energy facilities, is using AI systems to detect abnormal patterns in electricity generation and distribution, such as false data injection attacks or ransomware targeting supervisory control and data acquisition systems, and responding in real time to prevent outages or damage.

34 European Commission, "Tools and Techniques for Predictive IoT Security", Secure IoT, 31 Jul 2019. Available at: https://ec.europa.eu/research/participants/documents/downloadPublic?appId=PPGMS&documentIds=080166e5c6591672.

35 Attacker as a service ecosystem refers to the selling of necessary tools, infrastructure, and/or expertise to a less skilled malicious actor to carry out their desired attack.

36 European Union Agency for Cyber Security (ENISA), "Foresight Cyber Security Threats for 2030", ENISA, Mar 2024. Available at: https://www.enisa.europa.eu/sites/default/files/2024-11/Foresight%20Cybersecurity%20Threats%20For%202030%20Update%202024_0.pdf.

37 Polona Car with Tristan Marcelin, "Artificial intelligence and cyber security", Digital issues in focus, European Parliamentary Research Service, PE 762.292, Apr 2024. Available at: https://www.europarl.europa.eu/RegData/etudes/ATAG/2024/762292/EPRS_ATA%282024%29762292_EN.pdf.

38 Policy engines are software modules that evaluate system events or requests against predefined security or compliance rules and enforce appropriate actions (e.g. allow, block, or alert).

39 Security Orchestration, Automation, and Response (SOAR) playbooks are structured, automated workflows within SOAR platforms that guide the detection, investigation, and response to specific security incidents. For more information see Palo Alto Networks, "What Is SOAR?", available at: https://www.paloaltonetworks.com/cyberpedia/what-is-soar.

Industrial IoT and manufacturing sectors apply similar techniques to monitor connected machinery and sensors, automatically isolating compromised devices when malware infects industrial controllers (e.g. programmable logic controllers), or when stolen credentials are used to manipulate production lines. Transportation systems, such as rail networks and intelligent traffic management systems, are testing AI-driven responses to defend against threats posed by GPS spoofing and tampering with IoT sensors that control signaling and traffic flow. Even the healthcare industry is exploring the use of AI models and systems for monitoring networked medical devices, quickly identifying and mitigating abnormal behaviour, to protect against ransomware attacks or data exfiltration attempts.

The speed and sophistication of IoT attacks increasingly demand responses that are faster than what human intervention alone can provide. AI systems play a critical role in this regard, enabling automated containment, mitigation, and recovery strategies that keep pace with modern threats.[40]

One important capability is dynamic access control. Unlike static security rules, which remain fixed regardless of context, AI-driven systems continuously evaluate device behaviour and adjust security rules based on current conditions and observed behaviour. For example, if a device begins to act outside its normal pattern, the AI system can instantly quarantine it from the network to stop potential spread, downgrade its access privileges to ensure it has only the bare minimum rights, or require additional authentication such as a certificate re-handshake to revalidate its trustworthiness. These measures not only contain the immediate threat but also prevent lateral movement, thereby limiting the extent of a compromise and ensuring that attacks cannot easily propagate across the wider IoT ecosystem.

Beyond containment, AI systems also enable the emergence of autonomous, resilient network architectures. These architectures are designed not only to detect and isolate issues but to actively repair them in real time. For instance, if a device is compromised via a known vulnerability, the AI system can automatically deliver a patch or configuration update through a secure over-the-air process, reducing downtime and eliminating the risk of human error in manual updates. Similarly, in the event of a distributed denial-of-service attack targeting a specific network segment, AI systems can dynamically reroute legitimate traffic to unaffected pathways while filtering out malicious packets at the networks edge. This ensures that critical services remain operational even under active attack.[41]

These techniques can significantly reduce the mean time to respond to incidents, which is a crucial metric for resilience. Automated detection and recovery not only speed up the response cycle but also reduce the burden on human operators, allowing them to focus on strategic oversight and complex decision-making. In effect, AI models and systems transform IoT ecosystems into adaptive, self-protecting environments where threats are contained, systems are repaired, and operations continue with minimal disruption.[42]

Challenges of AI in IoT Security

While AI models and systems are powerful tools to enhance IoT security, they also introduce specific challenges that must be addressed:

- **Data Quality and Quantity:** AI models require vast amounts of high-quality, labelled data for effective training. Inaccurate, noisy, or biased IoT data can lead to poor model performance, resulting in false positives or negatives in threat detection. The complex and heterogeneous nature of IoT data makes data collection, cleaning, and normalisation a critical concern.[43]

40 Cyware, "SOAR and AI in Cyber Security: Reshaping your Security Operations", Securtiy Guide, accessed in Oct 2025. Available at: https://www.cyware.com/resources/security-guides/from-insight-to-action-how-ai-and-soar-are-reshaping-security-operations.

41 Ed. Keshav Kaushik et al., "Advanced Smart Computing Technologies in Cyber Security and Forensics", Taylor & Francis Group, 2022. Available at: https://www.scribd.com/document/696862281/Advanced-Smart-Computing-Technologies-in-Cyber Security-and-Forensics.

42 European Commission, "Tools and Techniques for Predictive IoT Security", Secure IoT, 31 Jul 2019. Available at: https://ec.europa.eu/research/participants/documents/downloadPublic?appId=PPGMS&documentIds=080166e5c6591672.

43 IoT CENTRAL, "Challenges of Artificial Intelligence and IoT Integration", accessed in Oct 2025. Available at: https://www.iotcentral.io/blog-all/challenges-of-artificial-intelligence-and-iot-integration.

- **Computational Overhead:** Running complex AI models on resource-constrained endpoint devices is often impractical. This limitation necessitates a distributed edge-cloud architecture where lightweight AI models operate on edge devices for initial processing, while heavier model training and inference occur in the cloud. Such architecture balances latency and resource use.
- **Adversarial AI:** Attackers may deploy their own AI techniques to evade detection by crafting adversarial inputs or poisoning training data, posing a significant threat to the robustness of AI systems. This necessitates continuous adversarial testing and monitoring.[44]
- **Explainability:** AI's "black box" nature may make it difficult to understand why certain activities are flagged as anomalous. This lack of transparency can hinder trust in AI systems and complicate investigation and incident response efforts.
- **Privacy and Compliance:** AI used in IoT systems often processes sensitive personal or operational data, raising concerns around data privacy and regulatory compliance. Ensuring end-to-end security and designing privacy-preserving learning methods – like federated learning, where models are trained locally on devices without sharing raw data with a central server – are essential to address these concerns.
- **Operationalisation Challenges:** Deploying, managing, and maintaining AI models at scale within complex IoT environments presents significant operational challenges, requiring robust monitoring, continuous model updates, and alignment with organisational policies.

These challenges emphasise the need for a holistic security strategy that integrates AI capabilities with strong data governance, robust architecture, privacy protections, and transparent operations to fully leverage AI systems for IoT security.

Better Practices for IoT Security

AI tools are transforming IoT security from a static, perimeter-based defence into a dynamic system. AI models and systems are essential for managing the scale and sophistication of modern threats, enabling organisations to transition from reactive defence mechanisms to predictive, adaptive, and autonomous response security operations across IoT systems. To facilitate this transition, organisations should consider the following:

- **Build Security into Devices and Fleets:** Follow the National Institute of Standards and Technology's IoT device capability catalogues and the 8259 series guidelines to establish security requirements for procurement and development. Key features include secure updates, authentication and authorisation, detailed logging, and configuration management. These foundational controls enhance the data quality available for AI detection systems and support safe automated responses.[45,46]
- **Harden the Data Pipeline:** Collect only high-quality and minimally necessary telemetry data, normalise it across different protocols, and maintain representative baseline datasets. In addition, track and monitor version datasets and models to manage data drift over time. This ensures AI models operate on reliable and consistent inputs, maintaining detection accuracy and system stability.

44 Mar Romero, "The Intersection of AI and IoT: Securing Connected Devices", NeuralTrust, 22 Apr 2025. Available at: https://neuraltrust.ai/blog/ai-iot-security-connected-devices.

45 Michael Fagan et al., "IoT Device Cyber Security Guidance for the Federal Government: IoT Device Cyber Security Requirement Catalog", NIST Special Publication 800-213A, U.S. Department of Commerce, Nov 2021. Available at: https://doi.org/10.6028/NIST.SP.800-213A.

46 Michael Faga et al., "Foundational Cyber Security Activities for IoT Device Manufacturers," NIST IR 8259, U.S. Department of Commerce, May 2020. Available at: https://csrc.nist.gov/pubs/ir/8259/final.

- **Adopt Privacy-Preserving Learning when Appropriate:** Leverage federated learning to enable collaborative model training across multiple sites without sharing raw data. Implement protections against federated learning-specific threats such as robust aggregation algorithms, differential privacy techniques, and the use of secure hardware enclaves to ensure data confidentiality and model integrity.
- **Integrate with Zero-Trust and Network Segmentation Architectures:** Use AI-generated insights to enforce identity-centric security policies, enable dynamic network segmentation, and apply least-privilege access controls. This is especially crucial for mixed IT and OT IoT environments facing emerging multi-vector threats.[47]
- **Defend the AI Model Itself**: Proactively monitor for data or model poisoning attacks and conduct adversarial robustness testing. Treat AI model updates with the same rigour as other critical infrastructure changes by gating changes to prevent unauthorised or harmful modifications.
- **Automate, but Keep Humans in the Loop**: Begin with supervised automated actions, such as device quarantine or traffic rate-limiting, while ensuring that rollback capabilities and comprehensive audit trails are in place. As AI explainability, accuracy, and operator confidence improves, gradually increase the level of automation autonomy maintaining a balance between efficiency and human oversight.

47 European Union Agency for Cyber Security (ENISA),, "ENISA Threat Landscape 2023", ENISA19 Oct 2023. Available at: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023.

The future of IoT systems will be characterised by increased automation and convergence with AI systems.

Future Trends: AI, Edge Computing, and IoT

The future landscape of IoT will be shaped by the deepening convergence of artificial intelligence, advanced networking, and cyber security. As IoT systems become more sophisticated and widespread, their security challenges will grow in complexity, necessitating equally advanced solutions. The next generation of IoT security will move beyond traditional perimeter defence toward intelligent, adaptive systems capable of predicting, preventing, and responding to threats in real-time. This evolution will be characterised by increased automation, with AI-driven security systems continuously monitoring device behaviour, network traffic, and data patterns to identify anomalies that may indicate a breach. The integration of AI into IoT security frameworks will enable more proactive defence mechanisms, fundamentally transforming how we protect connected devices and the critical infrastructure they support.

Edge AI: Bringing Intelligence to IoT Devices

Edge AI represents a fundamental shift in how we process IoT data by bringing computational capabilities closer to the data source. By deploying AI models directly on IoT devices or edge gateways, organisations can significantly reduce the latency associated with transmitting data to centralised cloud servers for processing. This localised approach enables real-time decision-making for time-sensitive applications such as autonomous vehicles, industrial robotics, and emergency response systems. Additionally, edge computing dramatically reduces bandwidth consumption by processing data locally and transmitting only relevant insights or aggregated information to the cloud. This not only lowers operational costs but also minimises network congestion, making IoT deployments more scalable and efficient, particularly in bandwidth-constrained environments.

Federated learning has emerged as a groundbreaking approach to privacy-preserving AI for IoT ecosystems. This technique enables model training across decentralised devices without exchanging raw data with a central server. Instead of collecting sensitive information from multiple endpoints, the system shares only model updates, weights and gradients, while keeping personal and proprietary data secure on local devices.

This approach addresses critical privacy concerns in applications such as healthcare monitoring, personal assistants, and industrial IoT where data confidentiality is paramount. By maintaining data locality while still achieving collective intelligence, federated learning enables organisations to comply with stringent data protection regulations like the European Union's General Data Protection Regulation and build trust with users concerned about privacy in an increasingly connected world.

Convergence of IoT, AI, and 5G/6G Networks

The integration of 5G and upcoming 6G networks with IoT and AI is revolutionising industrial automation by delivering unprecedented ultra-low latency communication. These advanced networks provide the reliable, high-speed connectivity necessary for mission-critical industrial applications that demand instantaneous response times. In smart manufacturing environments, this enables real-time control of automated systems, synchronous operation of collaborative robots, and seamless coordination between interconnected machines. The combination of IoT sensors, AI data processing, and 5G's ultra-reliable low-latency communication capability creates responsive production ecosystems where decisions are made in milliseconds, dramatically improving operational efficiency, enabling predictive maintenance, and reducing downtime in industrial settings.

The convergence of IoT, AI, and 5G/6G networks introduces sophisticated security capabilities through AI-driven network slicing. This technology allows the creation of multiple virtual networks on a shared physical infrastructure, each with customised security protocols and performance characteristics tailored to specific IoT applications.

AI algorithms continuously monitor network traffic patterns and device behaviour and offer threat intelligence to dynamically adjust security parameters for each network slice. This enables organisations to isolate critical infrastructure IoT devices on highly secure network segments while maintaining appropriate protection levels for less sensitive applications. AI-enhanced network slicing provides granular security controls, automatic threat containment, and adaptive security policies that respond to evolving threats in real-time, significantly strengthening the overall security posture of IoT ecosystems.

Regulatory and Ethical Considerations

As IoT technologies become increasingly pervasive, regulatory compliance with established security standards has become essential rather than optional. Frameworks such as the National Institute of Standards and Technology's Cyber Security for IoT program and the International Organization for Standardization and the International Electrotechnical Commission 27001 provide comprehensive guidelines for implementing robust security measures throughout the IoT device lifecycle. These standards address critical aspects including secure device identity management, data encryption, access control, and vulnerability management.

Compliance with these standards demonstrates an organisation's commitment to security best practices and helps mitigate legal and financial risks associated with data breaches. As governments worldwide introduce IoT-specific cyber security regulations, adherence to these internationally recognised standards will become increasingly important for market access, liability protection, and maintaining customer trust in an increasingly regulated and connected environment.

Depending on the level of autonomy, the integration of AI into IoT systems, which can range from semi-autonomous devices requiring occasional human input to fully autonomous systems operating independently, raises significant ethical considerations that extend beyond technical implementation. As these systems make increasingly consequential decisions without human intervention, ensuring ethical AI system behaviour becomes paramount. This involves addressing algorithmic bias that could lead to discriminatory outcomes, establishing clear accountability frameworks for AI-driven system decisions, and maintaining meaningful human oversight over autonomous systems.

Ethical AI implementation requires transparent decision-making processes, explainable outcomes, and mechanisms for auditing AI model and system behaviour. Organisations must develop comprehensive ethical guidelines that balance innovation with responsibility, ensuring that autonomous IoT systems operate fairly, transparently, and accountably while respecting human dignity and rights in increasingly autonomous environments.

VCDNP

Vienna Center for Disarmament and Non-Proliferation

The VCDNP is an international non-governmental organisation that promotes peace and security by conducting research, facilitating dialogue, and building capacity on nuclear non-proliferation and disarmament.

vcdnp.org

@VCDNP

info@vcdnp.org

VCDNP