Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

Security Challenges in AI-Enabled Robotics: Threats and Countermeasures

Kuldeep Kumar Kushwaha[1], Jaspreet Kaur [2]
School of Computer Science and Engineering, Lovely Professional University, Phagwara

Abstract: Artificial intelligence (AI) techniques like machine learning and computer vision have enabled transformative capabilities in robotics. However, integrating AI also introduces new vulnerabilities that could be exploited by adversaries to attack robots. This paper reviews emerging security threats and potential countermeasures for AI enabled robotics. Unique threats arise because robot behaviour depends heavily on training data which could be manipulated, and complex learned models whose logic is difficult to interpret. Key threats span the system lifecycle including data poisoning attacks, adversarial examples to fool perception, malware infecting software, supply chain tampering, and loss of control leading to safety risks. To address these concerns, robust training data governance, algorithmic hardening, cybersecurity best practices, safety engineering, and standards will be critical. Specific countermeasures include adversarial training to make models more robust, anomaly detection to catch unusual inputs, hardware roots of trust to verify integrity, fail-safes for safe operation, and regulation of ethical development practices. However, securing AI robotics remains challenging due to the evolving dynamics between attacks and defences, lack of verification methods for learned models, and fragmentation across vendor ecosystems. Safety, ethics, and security must be ingrained throughout the robotics development lifecycle. Collaboration between cybersecurity, AI safety, and engineering communities will be essential to develop holistic solutions. As robots proliferate into physical environments, advances in transparent and verifiable AI will help build trust. This paper provides a comprehensive survey of threats and countermeasures to guide future research towards securing our emerging AI enabled society.

***Keywords*:** Artificial Intelligence, Robotics, Security, AI enabled Devices, Cybersecurity

1. Introduction

The integration of artificial intelligence (AI) techniques like machine learning and computer vision has enabled remarkable advances in robotic capabilities. AI provides robots with the ability to perceive, learn, reason, and make decisions in complex real-world environments. This allows robots to perform tasks like navigation, object manipulation, and planning without requiring explicit programming of all behaviours. However, the reliance on AI also introduces new security vulnerabilities that could be exploited by malicious actors to attack, misuse, or take control of robots (Ahmad et al., 2021). As robots are increasingly deployed in sensitive applications like manufacturing, transportation, healthcare, and the home, it is critical to understand and address these emerging threats proactively.

This paper provides a comprehensive review of the security challenges introduced by AI robotics and promising techniques to mitigate them. We first provide background on how AI enables robotic capabilities and unique security considerations compared to traditional pre-programmed robots. Next, we categorize key threats facing AI robotics into five areas - data, algorithms, platforms, ecosystem, and humans. Example attacks are analysed in each category including data poisoning, adversarial examples, malware, privacy leaks, and loss of control. We then discuss potential countermeasures spanning across the system lifecycle including robust training data processes, algorithm hardening, cybersecurity best practices tailored for robots, safety engineering techniques, and regulation (Ahmad et al., 2021). Finally, open challenges are highlighted including the evolving dynamics between attacks and defences, lack of verification methods for complex learned models, and need for standardized benchmarks to systematically evaluate AI robotics security.

1.1 Rise of AI Robotics

AI techniques used in robotics include machine learning, neural networks, computer vision, reinforcement learning, planning, and knowledge representation (Belk, 2020). By learning from experience and data, robots can acquire capabilities like:

- Computer vision - Analyse camera images to understand environments, localize the robot, and detect objects using deep learning. Enables perception and navigation.
- Natural language - Understand commands provided in speech or text through natural language processing. Allows human-robot interaction.
- Reinforcement learning - Learn robot controllers and policies through trial and error interacting with the environment to maximize rewards.
- Motion planning - Plan optimal, collision-free robot motions and trajectories leveraging search algorithms. Can adapt reactively.

These learned capabilities allow robots to operate autonomously in complex, unstructured real-world environments and enable applications across manufacturing, transportation, healthcare, surveillance, and the home (Belk, 2020). The market for AI-enabled robots is expected to grow significantly in coming years. However, reliance on learning also introduces distinct security considerations.

1.2 Security Considerations in AI Robotics

Unlike pre-programmed deterministic robots, AI-enabled robots have unique security vulnerabilities:

- Dependence on training data - Performance depends heavily on large training datasets which could be manipulated by adversaries.
- Complex learned models - The reasoning inside models like deep neural networks can be difficult for humans to interpret. Logic is not hardcoded.
- Unpredictable emergent behaviour - Reinforcement learning policies lead to complex emerging behaviours that are challenging to verify in advance.
- Integration with IT systems - Network connectivity introduces cyber-attack surfaces. Cloud reliance increases potential targets.
- Interaction with humans - Safety critical if robots operate in close proximity with humans. Loss of control could lead to harm.

This necessitates a new security paradigm spanning the entire system lifecycle compared to traditional IT systems or robotics. In the next section, we discuss specific threats that exploit these inherent vulnerabilities (Brady, 1985).

Table 1 AI robotics security threats

THREAT CATEGORY	EXAMPLE THREATS
DATA THREATS	Data poisoning, data leakage
ALGORITHM THREATS	Adversarial examples, model extraction
PLATFORM THREATS	Malware, hardware trojans
ECOSYSTEM THREATS	Patching complexity, supply chain risks
HUMAN THREATS	Loss of control, privacy leaks

2. Background

Before analysing specific threats and countermeasures, we first provide background on how artificial intelligence enables robotic capabilities and unique security considerations compared to traditional robotics.

2.1 Artificial Intelligence for Robotics

Artificial intelligence (AI) is transforming robotics by providing the ability to perform complex tasks in unstructured real-world environments. AI approaches used in robotics include:

- Machine learning - Algorithms that can learn patterns from data in order to make predictions or decisions without explicit programming. Used for perception, planning, and control.
- Deep learning - Multi-layer neural networks trained on large datasets like images, video, and speech. Excels at pattern recognition problems like computer vision.
- Reinforcement learning - Agents learn behaviours through trial-and-error interactions with an environment to maximize rewards. Used for robot control policies (Brady, 1985).
- Computer vision - Algorithms for processing and analysing images or video to understand scenes and objects. Relies heavily on deep learning. Critical for environmental perception.
- Natural language processing - Understand or generate human language (text or speech). Enables natural verbal interaction and commands.
- Knowledge representation - Formally describe facts and rules related to objects, environments, tasks, and events. Supports robot reasoning and planning.

By applying these techniques, robots can autonomously perform complex tasks such as:

- Scene understanding - Analyse visual scenes to detect, categorize, and locate objects and positions even in cluttered environments. Enables navigation and manipulation.
- Motion planning - Plan collision-free trajectories and motions to achieve goals using search algorithms. Can incorporate dynamic constraints.
- Object manipulation - Grasp, lift, move, and place objects by coordinating perceptions with motions.
- Human-robot interaction - Understand natural language commands, gesture, gaze, and proceed with verbal dialog (Chen and Luca, 2021).
- Reinforcement learning control - Learn robot controllers and policies through experience interacting with the environment and people to maximize rewards.

AI is a key enabler for robot autonomy, allowing robots to handle ambiguity and adaptivity compared to pre-programmed control. However, this reliance on data and learned models also introduces new security considerations.

2.2 Unique Security Challenges in AI Robotics

While AI provides many benefits, it also creates distinct security risks compared to traditional robotic systems:

- Reliance on potentially vulnerable training data - Performance depends heavily on the quantity and quality of data used for training machine learning models. This data could be manipulated or poisoned by adversaries.
- Complex opaque learned models - The reasoning inside models like deep neural networks can be difficult for humans to interpret and verify. Logic is not hardcoded.
- Unpredictable emergent behaviour - Reinforcement learning based robot controllers lead to complex interactive behaviours that are challenging to fully predict or test in advance.
- Tight integration with IT systems - Network connectivity and reliance on cloud platforms expands the potential cyber-attack surface.
- Physical interaction with humans - Failures or loss of control pose direct safety risks if robots closely interact with humans in physical spaces.
- Ecosystem complexity - Supply chains, heterogeneous fleets, frequent updates, and fragmented standards create security management challenges.

These factors necessitate proactive security across the entire system lifecycle from data to algorithms to platforms and integration (Knasel, 1986). Security cannot be an afterthought in AI robotics - it must be ingrained into the design process. Next we elaborate on specific threats that exploit these inherent vulnerabilities.

3. AI Robotics Security Threats

- Data - Attacks aimed at manipulating or stealing the data used to train machine learning models.
- Algorithms - Attacks targeting the logic and outputs of AI algorithms and models.
- Platforms - Compromising the underlying computing hardware, software, and connectivity.

- Ecosystem - Challenges securing complex fleets of heterogeneous robots and integrations.
- Humans - Risks arising from interactions between robots and people.

Figure 1 AI robotics security threat categories

3.1 Data Threats

Data is a key enabler for AI algorithms (Knasel, 1986). The quantity and quality of data used to train machine learning models heavily influences their performance. However, data can also be the target of different attacks:

- Data poisoning - An adversary manipulates the training data used to develop AI models by injecting crafted malicious data points. For instance, an attacker could add incorrectly labeled images to the training set of an image classifier. This poisons the model, causing intentional failures during inference like misclassification. Data poisoning has been demonstrated experimentally for applications like autonomous driving.
- Data leakage - Attackers steal confidential training data through insider attacks or by exploiting vulnerabilities in data handling pipelines. Sensitive proprietary data provides adversaries valuable intelligence to mount further attacks and should be protected (Knasel, 1986).
- Data snooping - Monitoring inputs and outputs of models during inference can leak information about their structure and logic which could enable other attacks like model extraction. Data sanitization techniques like differential privacy help mitigate these risks.

Robust data governance and cybersecurity controls are critical across the data lifecycle to address these threats as data forms the foundation for AI algorithms.

3.2 Algorithm Threats

Given a trained AI model, adversaries can craft malicious inputs designed to induce intentional failures:

- Evasion attacks - Carefully perturb inputs to fool models and cause misclassifications without changing true semantics. For instance, adding small noise to images can fool deep neural network classifiers. This could trick computer vision for navigation or object recognition in robots.
- Model inversion - Reconstruct confidential training data from released models by using the model as an oracle to label queries. Allows adversaries to steal intellectual property or violate privacy.
- Model extraction - Obtain sensitive information about model structure, hyperparameters, or weights through querying APIs. Enables cloning proprietary models (Lozano-Perez, 1982).
- Model poisoning - Similar to data poisoning, directly manipulate model logic or parameters to degrade performance. Could be achieved through compromised insider access or exploits.
- Logic corruption - Tamper with model logic at runtime through faults induced on hardware or exploits targeting software. Could override programmed safe behaviours.

Defences aim to harden models against these threats through techniques like adversarial training, sandboxing, and formal verification (Lozano-Perez, 1982).

3.3 Platform Threats

The underlying computing platforms including hardware, software, and connectivity introduce cybersecurity risks:

- Malware infections - Bugs in complex software stacks assembled from many sources could enable malware that overrides robot behaviours. Worms could also spread between robots or into core IT systems.
- Rootkit attacks - Persistent malware embedded deep into robot software kernels, drivers, or firmware to maintain control while hiding from defences.
- Hardware trojans - Rogue malicious circuits implanted into integrated circuits that disable robots or leak data. Particularly concerning for electronics supply chain security.
- Remote cyber-attacks - Connectivity mechanisms like Wi-Fi or cellular networking expose vulnerabilities that could be exploited to disrupt robots remotely.
- Privacy leaks - Onboard sensors capture sensitive data about environments, users, or proprietary processes that may violate privacy or IP if accessed improperly.

Platform security requires a Défense-in-depth approach combining cybersecurity best practices with tailored techniques for robotics like hardware roots of trust.

3.4 Ecosystem Threats

Securing robot fleets and integrations with broader systems poses challenges:

- Patching complexity - Vast heterogeneous fleets running different software/hardware make patching vulnerabilities logistically difficult leading to long windows of exposure.
- Supply chain risks - Reliance on global supply chains increases risks of compromised or counterfeit hardware and software components (Okamoto, 2011).
- Ecosystem fragmentation - Lack of common security standards, architectures, and interfaces across the fragmented vendor ecosystem inhibits interoperability and cyber Défense.

A holistic systems view spanning fleets, supply chains, and lifecycles is imperative to manage ecosystem security.

3.5 Human Threats

The presence of humans working with robots raises additional safety and ethical concerns:

- Loss of control - Failures arising from other threats could lead to uncontrolled robot behaviour that injures nearby humans through collisions or falls. Strict safety constraints are necessary.
- Misuse - Intentionally abuse robot capabilities as weapons or tools for unethical, illegal, or dangerous ends ranging from petty theft to terrorism.
- Privacy - Improperly collect, expose, or use personal data about users captured by onboard sensors and cameras. Violates ethical principles and regulations.
- Social engineering - Humans may be susceptible to manipulations attacking organizations through robots. Social engineering training helps mitigate insider threats enabled by robots.

Governance frameworks encompassing regulation, codes of ethics, and engineering best practices are emerging to address societal impacts of AI robots (Okamoto, 2011).

This diverse range of threats necessitates a holistic security perspective spanning data, algorithms, platforms, ecosystems, and human interactions over the system lifecycle. Next we discuss promising techniques to counter these threats.

4. Security Countermeasures

In this section, we discuss promising techniques and best practices to counter the security threats facing AI-enabled robotics across the lifecycle.

4.1 Securing Training Data

Since performance depends heavily on training data quality (Tasioulas, 2019), organizations need robust data governance encompassing:

- Provenance tracking - Document the origin and collection process for data to ensure quality. Perform audits.

- Access control - Encrypt data and implement role-based access control for storage and pipelines. Prevent unauthorized access.
- Continuous monitoring - Monitor pipelines to detect anomalous data manipulations like poisoning.
- Configuration control - Manage data versions like code. Enable reproducibility and rollback.
- Privacy protections - Anonymize data and manage consent. Follow regulations like GDPR.

4.2 Hardening Algorithms

Various techniques make AI models more robust and secure:

- Adversarial training - Augment training data with adversarial examples to make models more resistant to evasion attacks.
- Anomaly detection - Detect inputs during inference that produce unusual model outputs not represented in training data as possible attacks.
- Sandboxing - Quarantine uncertain model predictions until validated to contain failures.
- Formal verification - Mathematically prove model properties within rigorously defined constraints. Early research for neural networks.
- Interpretability - Increase explainability of model logic to users to detect tampering.

4.3 Platform Cybersecurity

AI robotics requires a Défense-in-depth approach common in IT security:

- Network segmentation - Isolate robots from general networks with role-based access control.
- Behavioural monitoring - Detect anomalies in traffic and behaviours indicative of malware.
- Encryption - Secure communications between robots, operators, and cloud services.
- Hardware roots of trust - Use hardware protected encryption keys to verify integrity of firmware and software (Tasioulas, 2019).
- Principle of least privilege - Only provide software access to necessary robot functions and resources.

4.4 Securing Ecosystems

Improving ecosystem security requires:

- Information sharing - Promote sharing of threat intelligence between vendors and service providers.
- Secure remote updates - Standardize secure methods to patch vulnerabilities across fleets.
- Supply chain security - Leverage techniques like blockchain to verify integrity of hardware and software.
- Compliance testing - Validate robots meet security standards through independent audits.

4.5 Safe Development

Safety and ethics should be ingrained across the development lifecycle:

- Interpretability - Engineers must be able to explain model behaviours critical for safety.
- Fail-safes - Define safe fall-back operation if robots reach uncertain states. May require handover to human.
- Validation & verification - Rigorously test complex autonomous behaviours in simulations and constrained real settings.
- Regulations & standards - Develop consensus ethical principles and engineering best practices.

Table 2 Overview of AI robotics security countermeasures

Focus Area	Example Countermeasures
Data security	Provenance tracking
	Access control
	Privacy protections
Algorithm security	Adversarial training
	Anomaly detection
	Formal verification

Platform security	Network segmentation
	Behavioral monitoring
	Hardware roots of trust
Ecosystem security	Information sharing
	Secure updates
	Supply chain security
	Compliance testing
Human safety	Interpretability
	Failsafes
	Validation and verification
	Standards

5. Open Challenges

Table 3 Open challenges in securing AI robotics

Challenge	Description
Evolving attack-Défense dynamics	Continuous innovation needed to counter new attack techniques in an ongoing arms race between adversaries and defenders.
Verifying learned models	Testing alone is insufficient to guarantee safety. Formal verification methods for complex neural networks and learning systems remain immature.
Lack of benchmarks	Standardized benchmarks needed to systematically evaluate security of AI robots under realistic conditions.
Interdisciplinary expertise	Holistic security requires synthesizing specialized knowledge across domains like ML, cybersecurity, robotics, formal verification, safety engineering, and ethics.
Legacy systems integration	Integrating AI into legacy platforms not designed for security poses significant difficulties.

6. Conclusion

This paper provided a comprehensive overview of emerging security threats and promising countermeasures for AI-enabled robotics. The integration of artificial intelligence techniques like machine learning and computer vision enables remarkable new robotic capabilities but also introduces potential vulnerabilities that could be maliciously exploited. As robots are deployed into sensitive real-world environments, it is imperative that security is considered holistically across the entire system lifecycle (Vrontis et al., 2021).

We categorized key threats arising from AI robotics into five areas: data, algorithms, platforms, ecosystems, and humans. Specific threats span from data poisoning, adversarial examples, malware infections, supply chain

tampering, and loss of control causing safety incidents. To mitigate these concerns, techniques like robust data governance, algorithm hardening, cybersecurity best practices tailored for robots, safety-driven development methods, and regulation must be leveraged in combination. Promising countermeasures include adversarial training, behavioural monitoring, access control, fail-safes, and standards development.

However, securing AI-enabled robots remains challenging. The dynamics between attackers and defenders continue evolving rapidly, requiring ongoing innovation to counter emerging threats. Testing alone is insufficient to validate the safety of complex learned models whose logic is difficult to interpret. There is also a lack of standardized benchmarks to systematically evaluate AI robotics security under realistic conditions. Holistic solutions necessitate synthesizing specialized expertise across cybersecurity, machine learning, formal verification, robotics, ethics, and safety engineering. Further research and increased focus is imperative during development, well before any deployment.

As AI-enabled robotics continue proliferating into physical spaces and sensitive domains, the following conclusions and recommendations can help guide future progress:

1. Security must be ingrained early across the entire lifecycle - from data collection, through algorithm design, platform hardening, and integration testing. Bolting on security as an afterthought is dangerous.
2. Investment into explainable and verifiable AI techniques is critical, especially for safety-critical applications, to build trust and enable auditing.
3. Common standards for securing robotics platforms, applications, and data are urgently needed to improve Défense and manage complex fleets. These standards should encompass safety and ethics.
4. More extensive education on AI security threats and ethical obligations is required for practitioners and leadership across robotics, engineering, and AI. Security awareness must increase.
5. Proactive collaboration between cybersecurity, machine learning, and robotics experts from both industry and academia will be the key driver for impactful innovation to counter emerging threats.
6. Regulators may need to intervene as risks increase to ensure the societal benefits of AI robotics are realized while protecting the public. Policy and legal frameworks will help align development with social priorities.

The security challenges introduced by AI robotics should not deter progress but rather motivate holistic solutions to ensure this technology positively transforms society (Wang and Siau, 2019). With sustained research, standards, education, and cross-disciplinary teamwork, the world can safely unlock the remarkable potential of AI-enabled robotics across manufacturing, healthcare, transportation, the home, and many other domains. This paper synthesized the current state of knowledge to support that mission. There are always risks arising from powerful new technologies, but with wisdom and foresight, humanity can successfully navigate towards an AI-enabled future that benefits all.

Reference List

[1] Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y. and Chen, H. (2021). Artificial intelligence in sustainable energy industry: Status Quo, challenges and opportunities. *Journal of Cleaner Production*, [online] 289(289), p.125834. doi:https://doi.org/10.1016/j.jclepro.2021.125834.

[2] Belk, R. (2020). Ethical issues in service robotics and artificial intelligence. *The Service Industries Journal*, 41(13-14), pp.1–17. doi:https://doi.org/10.1080/02642069.2020.1727892.

[3] Brady, M. (1985). Artificial intelligence and robotics. *Artificial Intelligence*, 26(1), pp.79–121. doi:https://doi.org/10.1016/0004-3702(85)90013-x.

[4] Chen, Y. and Luca, G.D. (2021). Technologies Supporting Artificial Intelligence and Robotics Application Development. *Journal of Artificial Intelligence and Technology*, 1(1), pp.1–8. doi:https://doi.org/10.37965/jait.2020.0065.

[5] Knasel, T.M. (1986). Artificial intelligence in manufacturing: Forecasts for the use of artificial intelligence in the USA. *Robotics*, 2(4), pp.357–362. doi:https://doi.org/10.1016/0167-8493(86)90009-4.

[6] Lozano-Perez, T. (1982). Robotics. *Artificial Intelligence*, 19(2), pp.137–143. doi:https://doi.org/10.1016/0004-3702(82)90033-9.

[7] Okamoto, T. (2011). An artificial intelligence membrane to detect network intrusion. *Artificial Life and Robotics*, 16(1), pp.44–47. doi:https://doi.org/10.1007/s10015-011-0880-5.

[8] Tasioulas, J. (2019). *First Steps Towards an Ethics of Robots and Artificial Intelligence*. [online] papers.ssrn.com. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3413639.

[9] Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A. and Trichina, E. (2021). Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. *The International Journal of Human Resource Management*, 33(6), pp.1–30.

[10] Wang, W. and Siau, K. (2019). *Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity: A Review and Research Agenda*. [online] Journal of Database Management (JDM). Available at: https://www.igi-global.com/article/artificial-intelligence-machine-learning-automation-robotics-future-of-work-and-future-of-humanity/230295