HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

CHALLENGES OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN CYBERSECURITY AND ROBOTICS

Toshmurodova Gulsevar Bekmurod qizi
Student of the Faculty of Exact Sciences,
Department of Informatics, Artificial Intelligence and

Digital Technologies, Navoi State University

Doi: https://doi.org/10.5281/zenodo.21845716

Abstract: This article examines the major challenges associated with the application of artificial intelligence technologies in cybersecurity and robotics. Artificial intelligence has significantly improved threat detection, automated decision-making, system monitoring, and robotic control. However, its growing use has also introduced new technical, ethical, and security risks. In cybersecurity, AI systems may be exploited to create advanced malware, automate phishing attacks, bypass traditional protection mechanisms, and manipulate large volumes of sensitive data. At the same time, AI-based defence systems may produce inaccurate results, false alarms, or biased decisions because of poor-quality training data.

Key words: artificial intelligence, cybersecurity, robotics, cyber threats, autonomous systems, data security, algorithmic bias, privacy.

Introduction

Artificial intelligence has become one of the most influential technologies in the development of modern cybersecurity and robotics. Machine learning, deep learning, computer vision, natural language processing, and intelligent decision-making systems are increasingly used to analyse information, recognize patterns, automate complex processes, and respond to changing environments. In cybersecurity, artificial intelligence supports threat detection, network monitoring, malware analysis, user-behaviour assessment, and automated incident response. In robotics, AI enables machines to perceive their surroundings, plan movements, learn from experience, interact with humans, and perform tasks with different levels of autonomy.

The application of artificial intelligence offers significant advantages to cybersecurity systems. Traditional security tools often depend on predefined rules and previously identified attack signatures. Such methods may be insufficient when organizations face rapidly changing threats, unknown malware, large data volumes, and sophisticated attack strategies. AI-based systems can process extensive security data and identify unusual activities that may indicate an attack. They can also support security specialists by prioritizing alerts, detecting suspicious patterns, and accelerating responses to incidents. However, the relationship between artificial intelligence and cybersecurity is not entirely positive. AI technologies are dual-use tools because they may be employed for both defensive and malicious purposes. Cybersecurity professionals can use AI to identify threats, while attackers may use similar technologies to automate phishing, discover system vulnerabilities, modify malicious software, generate deceptive content, and avoid conventional security mechanisms. Consequently, artificial intelligence does not simply solve existing cybersecurity problems; it also creates new attack possibilities and increases the complexity of the digital threat environment.

One of the most serious problems is the vulnerability of AI models themselves. Machine-learning systems depend on training data, algorithms, software frameworks, computational

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

infrastructure, and user inputs. An attacker may target any of these elements. Data-poisoning attacks can introduce manipulated information into a training dataset and influence the behaviour of the resulting model. Evasion attacks can modify input data so that a trained system produces an incorrect classification or decision. Privacy attacks may reveal confidential information contained in or inferred from a model. Generative AI systems may additionally be exposed to misuse, manipulated instructions, harmful outputs, and attacks on their supporting tools and data sources

Literature Review

A significant contribution to the study of AI-related cybersecurity risks is the NIST report *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations*, prepared by Apostol Vassilev and his co-authors. The report presents a structured classification of attacks against machine-learning systems and establishes terminology for analysing the security of predictive and generative AI models [1]. It examines attacks according to the stage of the AI lifecycle, the attacker's goals, capabilities, knowledge, and the possible consequences for the system.

One of the principal contributions of the NIST report is its demonstration that AI security must be considered throughout the entire system lifecycle. Security problems may emerge during data collection, model training, testing, deployment, inference, updating, or integration with external tools. Protecting only the final model is therefore insufficient. Organizations must also secure training datasets, software dependencies, model files, user interfaces, computational infrastructure, and communication channels. The report identifies data poisoning as an important category of adversarial machine-learning attacks. In a poisoning attack, an adversary interferes with the data or processes used to train a model. The objective may be to reduce the general accuracy of the system or create a hidden behaviour that appears only when a specific input is presented. In cybersecurity applications, a poisoned model could incorrectly classify malicious activity as legitimate or produce unreliable threat assessments.

Evasion attacks occur after a model has been trained and deployed. An adversary modifies an input to influence the system's output without necessarily changing the model itself. This problem is especially relevant to cybersecurity tools that classify files, messages, network traffic, or user behaviour. If attackers understand how a detection system operates, they may modify malicious content to reduce the probability of detection.

Research Methodology

This study applies a qualitative descriptive methodology to investigate the main problems associated with artificial intelligence technologies in cybersecurity and robotics. The qualitative approach was selected because the risks of AI systems include not only measurable technical failures but also ethical, organizational, legal, privacy, and human-safety concerns.

The research is based on document analysis. Scientific publications, international cybersecurity reports, risk-management frameworks, and studies of AI-enabled robotic systems were examined. The methodological foundation includes the NIST Artificial Intelligence Risk Management Framework, which presents a structured approach to identifying, measuring, managing, and governing risks throughout the AI lifecycle [3]. The framework emphasizes that trustworthy AI should be valid, reliable, safe, secure, resilient, transparent, explainable, privacy-enhancing, and accountable.

The study also uses the findings of the European Union Agency for Cybersecurity concerning the relationship between artificial intelligence and cybersecurity. ENISA considers two interconnected

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

research areas: the use of AI to improve cybersecurity and the protection of AI systems from cyberattacks [4]. This distinction is important because artificial intelligence may function both as a defensive instrument and as a target or instrument of malicious activity. Cybersecurity problems in autonomous robotic systems were examined through examples from autonomous vehicles and uncrewed systems. The ENISA and Joint Research Centre report on autonomous driving was used to analyse attacks against sensors, machine-learning models, navigation systems, and safety-critical decision-making processes [5]. NIST research on uncrewed systems was used to examine how increasing autonomy, connectivity, software integration, and dependence on external data create new combinations of cybersecurity and AI risks [6].

The research material was divided into four analytical categories:

1. AI as a cybersecurity defence mechanism;
2. AI as a target and instrument of cyberattacks;
3. cybersecurity and safety risks in intelligent robotics;
4. ethical, legal, privacy, and human-control problems.

Several methods were applied during the analysis. The descriptive method was used to explain the main technological and security problems. Comparative analysis was employed to identify similarities and differences between cybersecurity systems and AI-enabled robots. Risk analysis was used to determine the potential consequences of system failures and attacks. The systems approach was applied to investigate the interaction between data, algorithms, sensors, communication networks, software, hardware, human operators, and organizational policies. The study also used thematic analysis. The selected sources were examined for recurring problems such as data poisoning, adversarial inputs, unauthorized access, model manipulation, privacy violations, biased decisions, weak authentication, sensor spoofing, insufficient explainability, and unclear responsibility.

The reliability of the analysis was supported by comparing findings from governmental, scientific, and cybersecurity sources. Instead of relying on a single technological perspective, the research considered both digital-security risks and the physical consequences of failures in robotic systems.

Analysis and Results

The analysis shows that artificial intelligence provides important advantages in cybersecurity. AI systems can process large volumes of network traffic, user activity, system logs, files, and security alerts more rapidly than manual analysis. Machine-learning models can identify patterns that may indicate malware, unauthorized access, fraud, phishing, or abnormal user behaviour.

Traditional cybersecurity tools often depend on predefined rules and known attack signatures. Although these methods remain important, they may fail to identify previously unknown or rapidly changing threats. AI-based systems can support anomaly detection by comparing current activity with patterns of normal behaviour.

For example, an AI cybersecurity system may detect an account that suddenly connects from an unusual location, downloads an abnormal amount of data, or attempts to access protected resources. Such behaviour does not automatically prove that an attack is taking place, but it can help security specialists prioritize an investigation.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

AI also supports automated incident response. A security platform may temporarily block suspicious traffic, isolate a compromised device, request additional authentication, or notify cybersecurity personnel. These functions can reduce the time between the beginning of an attack and the organizational response. However, the results demonstrate that AI should support rather than completely replace human cybersecurity specialists. Automated systems may produce false positives by classifying legitimate actions as malicious. They may also generate false negatives by failing to recognize an actual attack. Human review remains necessary when decisions may affect important systems, personal data, financial operations, or organizational continuity.

The findings reveal that artificial intelligence has a dual-use character. The same technologies that assist defenders may also improve the capabilities of attackers. AI can be used to generate convincing phishing messages, imitate writing styles, automate social-engineering attacks, identify software weaknesses, and modify malicious content.

Generative AI may allow attackers to produce large numbers of personalized messages. Traditional phishing messages often contain grammatical errors or general content. AI-generated messages can be adapted to a specific organization, profession, event, or individual, making deception more convincing. AI tools may also assist in analysing publicly available information about an organization. An attacker could use this information to identify employees, technologies, suppliers, or business processes and then create a more targeted attack. Therefore, AI may increase not only the speed but also the personalization of malicious activity. Another problem is AI-assisted malware modification. Attackers may attempt to change malicious code or communication patterns so that traditional detection tools do not recognize them. This creates continuous competition between AI-based defence systems and AI-assisted attacks. The analysis indicates that organizations should not treat AI-generated threats as completely separate from traditional cybersecurity. Phishing, malware, unauthorized access, and data theft remain familiar threats, but AI can make them faster, more adaptable, and more difficult to detect.

One of the most serious findings concerns attacks directed against AI systems themselves. Artificial intelligence depends on data, models, software libraries, computational resources, and interfaces. A weakness in any of these components may influence the reliability of the final decision. Data poisoning occurs when incorrect, misleading, or malicious data are introduced into a training dataset. If a cybersecurity model learns from manipulated data, it may classify harmful behaviour as safe or legitimate activity as dangerous.

For example, an attacker may repeatedly introduce specially designed network activity into a system so that the model gradually accepts the pattern as normal. When the actual attack occurs, the system may fail to recognize it. Adversarial or evasion attacks occur when inputs are deliberately modified to produce an incorrect output. The modification may be difficult for a human to notice but sufficient to influence a machine-learning model. Such attacks are particularly dangerous in computer vision, malware detection, facial recognition, autonomous navigation, and biometric authentication.

Model theft and information-extraction attacks create additional concerns. An attacker may attempt to reproduce the behaviour of a protected AI model or infer confidential information from its responses. This is especially serious when the system has been trained on personal, medical, financial, or organizational data.

The results show that AI security requires protection throughout the system lifecycle. It is not sufficient to secure only the final application. Organizations must protect data collection, model

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

training, testing, deployment, updating, access permissions, software dependencies, and communication interfaces [3; 4]. The quality of AI decisions depends heavily on the quality of training data. Incomplete, outdated, inaccurate, or unrepresentative datasets may create unreliable results.

Table 1.

Major AI-Related Challenges in Cybersecurity and Robotics

Area	Main Challenge	Potential Consequences	Recommended Measures
Cybersecurity	AI-assisted phishing and malware	More personalized attacks, faster malware modification, and increased difficulty of detection	Continuous threat monitoring and combination of AI-based and traditional security mechanisms
AI Models	Data poisoning and adversarial attacks	Incorrect classification, false decisions, and failure to detect malicious activity	Protection of training data, model testing, lifecycle security, and continuous monitoring
Data Quality	Incomplete, biased, or outdated datasets	False positives, false negatives, biased decisions, and reduced model reliability	Regular dataset evaluation, updating, and model-performance assessment
Robotics	Sensor spoofing and manipulated inputs	Incorrect object recognition, unsafe navigation, collisions, or physical damage	Sensor validation, secure communication, and continuous environmental monitoring
Robotic Control Systems	Unauthorized access to control mechanisms	Manipulation of robot speed, direction, force, or assigned tasks	Strong authentication, encryption, access restrictions, and emergency-stop mechanisms
Privacy	Collection and misuse of personal data	Unauthorized disclosure of images, voices, locations, routines, and user information	Transparent data policies, controlled access, and secure data storage
Human Oversight	Excessive dependence on autonomous decisions	Unexplained errors, unsafe actions, and unclear legal responsibility	Human review, explainable AI, emergency controls, and clearly defined responsibility

As shown in Table 1, the challenges associated with artificial intelligence extend beyond conventional cybersecurity threats. The comparative analysis demonstrates that AI-related risks can affect data, algorithms, communication networks, sensors, control mechanisms, privacy, and physical safety. While cybersecurity systems are primarily exposed to risks such as phishing, malware, data poisoning, and adversarial attacks, AI-enabled robotic systems face additional physical risks resulting from sensor manipulation, unauthorized control, and navigation failures. Therefore, effective AI risk management requires a comprehensive approach combining technical protection, continuous monitoring, data-quality control, explainability, and meaningful human oversight.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

In cybersecurity, biased data may cause the system to concentrate on familiar attack patterns while ignoring new forms of malicious behaviour. A model trained mainly on one network environment may not perform reliably in another organization with different users, devices, or communication patterns. In robotics, poor data may cause an AI system to incorrectly identify objects, people, obstacles, or environmental conditions. A service robot trained in controlled laboratory conditions may behave differently in crowded, noisy, or unpredictable real-world environments.

Bias may also affect human-related decisions. Facial recognition, voice recognition, and behavioural analysis systems may perform unevenly across different groups or environmental conditions. Such errors can create privacy, fairness, safety, and legal problems.

The findings demonstrate that data quality should be evaluated continuously rather than only during initial system development. New environments, user behaviour, technologies, and attack methods may reduce the accuracy of an existing model. Therefore, datasets and model performance must be reviewed and updated regularly.

The analysis reveals that AI-enabled robots face more complex risks than ordinary software because they connect digital decisions with physical actions. A compromised information system may result in data loss, but a compromised robot may also move incorrectly, damage equipment, interrupt production, or harm people.

An intelligent robot normally depends on perception, planning, communication, and control components. Cameras, microphones, GPS receivers, LiDAR, motion sensors, and other devices collect environmental information. AI models interpret the information and select an action. Control mechanisms then convert the decision into physical movement.

A failure in the perception stage can influence the entire system. If a camera or sensor provides incorrect data, the robot may create an inaccurate model of its surroundings. It may fail to recognize an obstacle, person, traffic sign, or dangerous area.

Sensor spoofing is a significant threat. An attacker may provide false GPS signals, manipulate visual objects, interfere with wireless communication, or create misleading sensor inputs. The ENISA and JRC analysis of autonomous driving demonstrates that small physical modifications, such as altered road signs or markings, may cause an AI perception system to classify its environment incorrectly [5]. Navigation and planning systems are also vulnerable. If an attacker manipulates maps, location information, or route-planning data, a robot may enter a prohibited area or select an unsafe path. In industrial environments, such behaviour may interrupt production or create a collision risk.

The control component represents a particularly critical attack surface because it directly manages motors, tools, and other physical mechanisms. Unauthorized commands could change a robot's speed, direction, force, or task. Therefore, control systems require strong authentication, encrypted communication, access restrictions, and emergency-stop mechanisms.

Modern robots commonly depend on wireless networks, cloud platforms, remote-control applications, software updates, third-party libraries, and external data services. These connections improve functionality but also expand the attack surface. Weak passwords, insecure network protocols, outdated software, and incorrectly configured interfaces may allow unauthorized access. An attacker may intercept data, manipulate instructions, disable the device, or take partial control of robotic functions. Software supply-chain security is another important concern. A robotic system may include

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

components developed by several companies. A vulnerability in a library, sensor, operating system, cloud service, or update mechanism can affect the complete system.

NIST research indicates that increasingly connected and autonomous uncrewed systems create risk-management problems extending beyond the individual robot. Their operation may depend on control stations, planning software, maintenance platforms, external databases, emergency systems, and communication with other devices [6].

AI-enabled robots may collect large amounts of personal information. Domestic, educational, healthcare, security, and service robots can record images, voices, movements, locations, routines, and user preferences. This information may be necessary for providing personalized services, but it may also be accessed, reused, or shared without sufficient protection. Privacy risks become particularly serious when robots operate in homes, hospitals, schools, or other sensitive environments.

Human trust also creates a security problem. People may communicate openly with robots that use natural language, recognize faces, or display social behaviour. Users may reveal confidential information because they perceive the system as helpful or trustworthy. A compromised robot could exploit this trust. It might request unnecessary information, provide false instructions, or influence user behaviour. Consequently, users should be clearly informed about what data the robot collects, how the data are processed, where they are stored, and who can access them.

The analysis identifies limited explainability as another major challenge. Complex AI models may produce decisions without providing an understandable explanation. In cybersecurity, specialists need to understand why an event has been classified as malicious. Without explanation, it becomes difficult to confirm an alert, correct an error, or justify a response.

In robotics, explainability is especially important after an accident or unexpected movement. Developers and operators must be able to determine whether the problem resulted from the data, sensor, model, software, communication network, hardware, or human instruction.

Legal responsibility may also be unclear. When an autonomous robot causes damage, possible responsibility may involve the manufacturer, software developer, data provider, system integrator, owner, operator, or organization using the system.

The findings support the continued use of meaningful human oversight. High-risk decisions should not depend entirely on automated output. Humans should be able to review decisions, interrupt unsafe actions, activate emergency controls, and recover the system from failure.

References:

1. Vassilev, A., Oprea, A., Fordyce, A., Anderson, H., Davies, X., & Hamin, M. (2025). *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations*. NIST AI 100-2 E2025. National Institute of Standards and Technology.
2. Neupane, S., Mitra, S., Fernandez, I. A., Saha, S., Mittal, S., Chen, J., Pillai, N., & Rahimi, S. (2024). Security considerations in AI-robotics: A survey of current methods, challenges, and opportunities. *IEEE Access, 12*, 22072–22097.
3. Tabassi, E. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. National Institute of Standards and Technology.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

4. European Union Agency for Cybersecurity. (2023). *Artificial Intelligence and Cybersecurity Research*. Athens: ENISA.

5. European Union Agency for Cybersecurity & Joint Research Centre. (2021). *Cybersecurity Challenges in the Uptake of Artificial Intelligence in Autonomous Driving*. European Union Agency for Cybersecurity.

6. Sheh, R., Geappen, K., & Harriss, D. (2024). *Autonomous Cybersecurity and AI Risk Management for Uncrewed Systems: Challenges and Opportunities Using the NIST Frameworks*. AUVSI XPONENTIAL 2024.