Cyber Vulnerabilities and Risks of AI Technologies in Space Applications

Paola Breda, Rada Markova, Adam Abdin, Devanshu Jha, Antonio Carlo, Nebile Pelin Mantı

IAC-22-D5.4.1.x70380

Cyber Vulnerabilities and Risks of AI Technologies in Space Applications

Paola Breda[a]*, Rada Markova[b], Adam F. Abdin[c], Devanshu Jha[b], Antonio Carlo[d], Nebile Pelin Mantı[e]

[a] *HyImpulse Technologies GmbH, Neuenstadt am Kocher, Germany,* breda@hyimpulse.de
[b] *Space Generation Advisory Council (SGAC), Vienna, Austria,* rada.markova92@gmail.com
[c] *Université Paris-Saclay, CentraleSupélec, Laboratoire Génie Industriel, 91190, Gif-sur-Yvette, France,* adam.abdin@centralesupelec.fr
[d] *Tallinn University of Technology, Tallin, Estonia,* ancarl@taltech.ee
[e] *PIL Department, Istanbul University, Faculty of Law, Istanbul, Turkey,* np_manti@yahoo.com
* Corresponding Author

Abstract

Artificial Intelligence (AI) is becoming a key technology for space applications. Recently, AI has come into extensive use in spacecraft operations, for example to support highly efficient operations of satellite constellations. This ranges in applications from relative positioning, Earth Observation, autonomous navigation, and end-of-life management, among others. While the importance of AI is rising for new space assets, AI is vulnerable to cyber threats, and AI cyber security is becoming an important aspect of space safety and operational security. This work aims to identify the vulnerabilities that AI systems may introduce to space assets and to analyse the potential operational threats and effective technological and regulatory mitigation measures. Towards this goal, the paper first examines and differentiates between vulnerabilities in legacy space systems, and those that are particularly related to AI technologies. The analysis covers the definition of AI technology as well as a detailed discussion about its current use in space related applications. Secondly, a comparison between prevailing cyber-attacks in space and cyber-attacks targeting AI technologies is made. Based on this assessment, the paper recommends prevention and mitigation measures that are contingent on cyber resilience of space operations focusing on AI-based space applications.

Keywords: Artificial Intelligence, Space Applications, Cyber-vulnerabilities, Prevention, Mitigation

Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
CIA	Confidentiality, Integrity, Availability
CCSDS	Consultative Committee for Space Data Systems
COTS	Components Off the Shelf
FCC	Federal Communications Commission
HW	Hardware
IoT	Internet of Things
ITU	International Telecommunication Union
INCOSE	International Council on Systems Engineering
LEO	Low Earth Orbit
ML	Machine Learning
NIST	National Institute of Standards and Technology
NSR	Northern Sky Research Report
PNT	Position Navigation and Timing
SW	Software
TT&C	Telemetry, Tracking and Command

1. Introduction

The human exploration of space and the increasing interest in the small satellite constellation deployment for Low Earth Orbit (LEO) are only two examples of how space missions can cover a broad variety of technologies, mission scopes and level of complexity. From control of landers and rovers to the increasing need of performing collision avoidance manoeuvres in space without support from ground, new challenges call for radical advances in several areas of space engineering. The design of intelligent space assets is one of such challenges.

Artificial intelligence (AI) is pervasive in an abundant number of applications throughout a broad spectrum of sectors and industries. The increasing interest of the entire aerospace community towards AI is strong in the NewSpace economy [1]. The European Investment Bank reveals that AI is a technological trend particularly suitable for satellite services and ground equipment, as well as a tool for processing big data or imaging [2]. AI can help to significantly reduce the operational cost of satellite operations, for example by optimising the satellite trajectory, or by augmenting its space situational awareness [2].

Nonetheless, the use of AI in the field of space raises open questions and challenges. As an example, to

guarantee the safety of the operations across the mission phases, typically, a mission work package on risk assessment is included in space projects. Its goal is to identify and reduce risks and protect space systems during normal operations. The use of AI technologies, however, can introduce additional vulnerabilities at the system level, as will be thoroughly discussed in this work. This article deals in particular with cyber-vulnerabilities generated by the use of AI technologies in space.

Although cybersecurity threats on space assets are part of a risk scenario known by most space actors [3], a specific assessment for AI-driven space systems is missing in the literature. To address this need, there is an on-going process of transferring the cybersecurity expertise from the information technology industry to the space industry. Additionally, engineering standards can be tailored to the space industry, as demonstrated by the International Council on Systems Engineering (INCOSE) [4]. However, the lack of technical standards specific to AI space applications makes it extremely hard to define relevant safety and security requirements. While national and regional space agencies are drafting guidelines for space cybersecurity, these do not specifically regulate or cover the use or risks of AI technologies. One example is given by the National Institute of Standards and Technologies from the US Department of Commerce (NIST), which proposed standards for risk assessment in space missions, covers the cybersecurity aspect [3], but does not mention AI-systems.

On this premise, this work aims, in the first place, to identify the additional vulnerabilities introduced by AI systems for space missions compared to legacy systems. Then, new scenarios covering cyberattacks on AI systems in space are presented to complete the study. Finally, a regulatory framework applicable to AI technologies in space applications is proposed. The purpose of such a framework is to act as a starting point for the mission planners to derive high-level requirements to ensure cyber resilience of AI-driven space systems and mitigate the consequences of cyber-attacks conducted against such systems.

2. Space mission vulnerabilities in legacy systems

In this section, the paper discusses cyber vulnerabilities of legacy space systems. The discussion is contingent on consideration of common cyber threats and operational segments of space missions.

2.1 Cyber threats

To define the context of cyber-vulnerabilities and attacks to space systems, it is useful to introduce the concept of threat originally presented for information systems [5], and how it differs from a risk [6]. A threat consists of four components, namely agents or sources, targets, actions, and consequences as reported schematically in Fig. 1.

A threat represents the potential for a threat agent to cause loss or damage to an information system (a space system) [6]. Note that this definition applies to both legacy and AI-based space technologies addressed in this work. A threat agent or source is a group or entity triggering an action causing damage to a threat target. According to the NIST classification [3] agents can be adversarial, insider, environmental, or structural. A threat agent usually exploits the vulnerabilities of the space system to perform an attack.

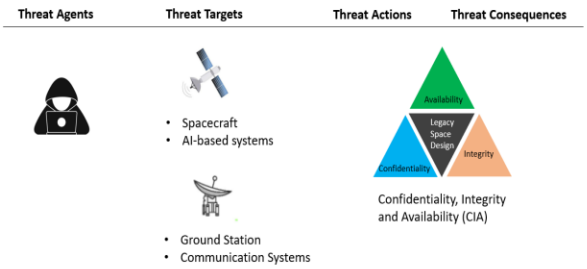

Fig. 1. Cyber threats in space (adapted from [6])

The threat action represents the event of attacking a target subjected to a threat, while the threat consequences can be identified in the adverse impact of the attack on the system and mission level. Threat actions and consequences can be linked using the Confidentiality, Integrity and Availability (CIA) plot (adapted from the Northern Sky Research - NSR Report [7]). The triangle shows the three main objectives for data handling in the security architecture. A risk, on the other hand, represents both the probability that a system will be the threat target and the magnitude of the harm caused.

The assessment and identification of cyber threats to space missions is a crucial task for mission planners. Guidelines for vulnerability and risk assessments for AI-systems in space applications are scarce in the literature. An example of applicability for space mission planning is provided by the Consultative Committee for Space Data Systems (CCSDS) Recommended Standards [6] which is used as reference for the discussion in this paper.

2.2 Operational segments of a space mission

Space missions are, generally, characterised according to three operational components: the space segment, the ground segment and the user segment, as illustrated in Fig. 2. Note that a fourth segment, the link segment, can be found in the literature when dealing with cybersecurity. The link segment is described by the crosslink, uplink and downlink arrows in Fig. 2. The assessment of cyber vulnerabilities of a space mission, typically, focuses on the first two components.

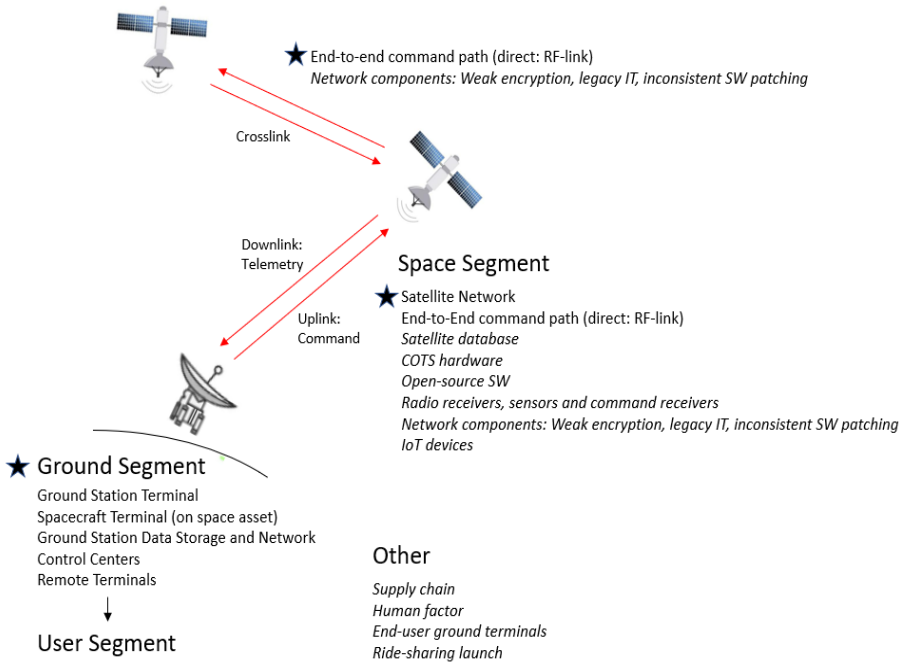

Fig. 2 Typical satellite mission - ground, space, and user segments, with their main cyber-vulnerabilities.

For the ground segment, ground stations provide telemetry and command to the spacecraft (shown by the red arrows in Fig. 2). Ground networks ensure the connection between ground elements. Control centres allow the management of space operations while remote terminals act as an interface to retrieve transmitted information.

The space segment consists of the payload and the spacecraft bus [8]. The direction of communications between spacecraft and Earth is illustrated by the red arrows in Fig. 2, and they include:

- downlink command for telemetry and tracking (space to ground)
- uplink command for operational manoeuvres (ground to space)
- crosslink for communications with other satellites (space to space)

2.3 Cyber-vulnerabilities in space

The identification of cyber vulnerabilities in space for this work is limited to space operations in LEO, as it is the most complex network of communication between ground-to-satellite and satellite-to-satellite. The use of constellations requires the interaction of hundreds or thousands of satellites in LEO with tens or thousands of ground network nodes, or millions of antennas [7].

The ground segment, as shown in Fig. 2, is responsible for collecting and distributing the mission data, as the most valuable asset of a space mission [8]. A generic definition of end-user is to be intended here because the user segment shown in Fig.2 is not analysed in this work. The ground infrastructure is the most vulnerable among the mission segments and it is used frequently as a threat target. In fact, access points to a ground segment would allow agents with legitimate access to control the infrastructure.

NASA reports that the most accessible portion of a spacecraft is the end-to-end command path [8] (red arrows in Fig. 2). The command path can be accessed via direct RF-link, subversion of the command authority on ground or subversion of space/ground networks. These items are represented by a star in Fig. 2.

In a spacecraft, the most vulnerable sub-systems to cyberattacks are Telemetry, Tracking and Command (TT&C), electronics/avionics, and On-Board Data Handling (OBDH) as they all act as access points for threat agents. In general, a threat agent could potentially exploit the vulnerabilities of any sub-components of a spacecraft or ground system to take control of the overall system.

Depending on whether the spacecraft was manufactured following a legacy design or a NewSpace philosophy, characteristic vulnerabilities can be recognised. For example, old network components that do not conform to current cybersecurity standards or trends can be found in legacy spacecraft because 15-20 years ago cybersecurity was not yet a major concern. Satellites developed in more recent years implement data encryption to protect data confidentiality in transmissions or to prevent

unauthorised access. In legacy systems, weak cryptographic units could represent a vulnerability in the end-to-end command link (weak encryption in Fig. 2). Nowadays, technical standards are proposed based on the mission scope. For example, NASA requires propulsive spacecrafts within 2 million kilometres from Earth to protect command uplink with encryption compliant with Level 1 of FIPS 140-3 [9]. The Federal Communications Commission (FCC) has considered requiring encryption on the TT&C communications and mission data for propulsive spacecraft, however no specific requirements are published to-date. The technical standard from NASA STD-1006 [10] covers the protection of the Position Navigation and Timing (PNT) to reduce the consequences of a loss or temporary interference of GNSS signals, as this attack profile is becoming more frequent.

Cyber-vulnerabilities are also associated with the supply chain. For instance, hardware and software can be maliciously modified undetected, so that the vulnerability of the system is presented at a later stage during the operation phase.

Products developed in the NewSpace economy introduce new vulnerabilities compared to legacy technologies [7] expanding the attack surface (see Table 1). Examples of such products include COTS hardware, open-source software, IoT devices on-board and infrastructure as a service (aaS).

The employment of AI-based space applications adds further vulnerabilities to those presented above, most of which are intrinsic to AI technology itself.

Table 1. Cyber vulnerabilities for NewSpace

Vulnerability	Explanation
COTS hardware (HW); Open Source software (SW) [7]	Non-compliant with cybersecurity recommended practices; dependency on vendor trustworthiness
Ridesharing: multi-customer; multi-payload [8]	The number of third-party actors which require authentication on the system increases
As-a-Service infrastructure (aaS)	Ground systems based on cloud
Internet of Things (IoT) devices	Increasing entry points, major risk of backdoor holes

3. AI for space missions

Despite the widespread adoption of the term AI, there is no generally agreed and established definition of AI in the literature, but its description is context specific. Regardless of the specific sector of its application, AI exhibits common characteristics which allow it to outline certain of its most distinctive features.

3.1 The definition of AI

AI encompasses a set of methods capable of autonomous decision making without human intervention [11]. Such methods are capable of adaptation to complex environments, and to previously unseen circumstances [12]. Most applications are related to the sub-field of Machine Learning (ML), which combines algorithms, statistics, and optimization theory. ML primarily aims at using information extracted from the environment to find patterns and structures that help in decision making in similar environments. To this extent, AI is intended to function as a proof of knowledge. The outcome of a decision-making process in real life may vary according to the data type used to train the algorithm (e.g., supervised, unsupervised, reinforcement learning).

AI is often characterised in the literature by data and training processes. Nonetheless, the field of AI does not cover a group of training algorithms only, but it also deals with different layouts of network architectures and different models applicable during the learning phase. As an example, Artificial Neural Networks (ANNs), which are among the most successful implementations of ML [13], have the properties of a network-like structure. Other AI elements include architecture related requirements given by the specific algorithms, i.e., AI software systems which are built on the underlying hardware [14].

In connection with the above-mentioned architecture requirements, oftentimes AI is associated with decision-making processes [15] which are similar, but not identical, to the decision-making process performed in the context of space autonomy. Although AI is closely related to space autonomy, which consists in one of its central features, a confusion between the two concepts is discouraged since they have different meanings. Notably, space autonomy characterises a set of functions of the system while AI designates specific techniques by which they are implemented, enabled, or enhanced. On the other hand, AI shares some of its central characteristics with space autonomy. For instance, space autonomy depicts a behaviour-oriented ability of a system where certain functions are allocated to a robot (or system) whereas other to a human [16]. AI is developed and deployed in the same machine – human paradigm, even more in the space sector where space missions are strongly dependent on the ground segment and flight engineers who monitor the telemetry data sent back to Earth [17].

In the inference stage, where the model is used to make decisions based on unseen input data [18], the output could be potentially wrong, if the input data was not represented by the training dataset. The subsequent behaviour may be unexpected, even more when the system operates in an uncertain environment, such as outer space. Non-determinism, thus, is another feature of AI.

Non-deterministic behaviour needs to be differentiated from a behaviour-based system, which reacts to a changing environment in an instinctual way [19]. The two concepts may meet as a behaviour-based system could also exhibit a non-deterministic behaviour [19]. This is the case of distributed artificial intelligence where AI is located in more than one part of the system (or more than one agent composing one system).

Other than space autonomy, AI also serves to enhance optimisation in the field of space system design. Combining different AI abilities/technologies at system-level is rather rare in practice due to little research on how to integrate modules from different areas of AI [20]. As long as AI is part of one common system, it is configured to operate together with non-AI elements – hence, the need to consider AI from a system perspective.

AI systems are inherently more complex than classical decision-making systems. An AI learning model is embedded within a software architecture and constitutes a complex interaction between an algorithm and a dataset to establish the decision-making system. Such systems may introduce additional sources of vulnerability to the system architecture, as presented in the next section.

3.2 Vulnerabilities of AI

Vulnerabilities in AI systems refer to instances where the model performs well (no errors in the initial implementation for the desired task) but is susceptible to malfunction in specific conditions, as a result of unidentified performance occurring during execution or provoked intentionally by an adversary [11].

An overview of the vulnerabilities of AI systems is discussed in this section, with the aim of understanding how such vulnerabilities can be exploited when applied to space systems. Table 2 provides a summary of these vulnerabilities and their description.

Table 2. Vulnerabilities of AI systems in general

AI vulnerability	Description
Low explainability	Black-box algorithms
Input data	During training phase
Data poisoning	During deployment phase
Mathematical models	During development phase
SW/HW interfaces	Cyber-physical system

The lack of explainability of many AI algorithms and outputs is one of the main issues underlying the vulnerability of these systems. Black-box AI algorithms that are not transparent nor easy to explain offer several opportunities for natural or adversarial malfunctions to go unnoticed. Improving the transparency and explainability of AI systems improves the user trust [21]. However, this can be problematic when plausible explanations are provided for outputs that are incorrect [22].

Input data is a key vulnerability to most AI systems. Attacks on data can occur during the training phase. ML models can be vulnerable to perturbations in the input data, leading the learning model to misinterpret or misclassify the data [23]. An attacker who wishes to exploit this vulnerability does not need to completely transform the input, and in many cases, does not need to be deeply familiar with the exact functioning of the model [24]. Similarly, data poisoning can occur after the model is properly trained and deployed. In this case, imperceptible manipulation of the data may lead to a fundamental malfunctioning of the system.

Moreover, ML models employ mathematical procedures that may have inherent weaknesses. In the development phase, an adversary with malicious intentions may exploit a specific model architecture known to be susceptible to various errors, such as unreliable and noisy outputs. Models are also subjected to replication attacks which allow an adversary to reverse-engineer a model. Such procedures can be used to understand and design a malfunction for the original system, or to steal intellectual property [25].

3.3 AI in space applications

Due to the improvements in hardware and computational capabilities of the recent decades, AI systems can carry better sensors and run sophisticated algorithms in a compact form factor [26]. Several works studied the applications of AI in the space sector [27], covering AI applications for remote sensing [28], on-board spacecraft communication [29], situational awareness [30], autonomous planning and scheduling [31], and guidance and control [16]. Three main areas of application for AI in the space sector are discussed here, particularly, those related to remote sensing, autonomous navigation and spacecraft health monitoring. Applications for satellite communications are still rare and, therefore, will not be covered in this work.

3.3.1 Data processing and remote sensing

Remote sensing broadly encompasses the objective of feature identification of distant objects using electromagnetic radiation. The amount of data collected, and their variety and complexity made it impossible to be processed by a human operator so that they are increasingly relying on AI algorithms. Satellites can employ on-board deep-learning algorithms for pre-processing of the sensory data to reduce the amount of data transmitted to ground segment [32] or to directly transmit post-processed information.

Remote sensing for features identification and extraction can be regarded as a primary step for proper classification of remote sensing imagery. In many cases, observations made about an object are significantly enhanced if this information comes from a variety of sources, sensors, processes and measuring states as well

as at different spatial and temporal resolution. Remote sensing for data and sensor fusion couples the information from different sources, including in-situ observations and numerical models to improve the quality of the information extracted. Combining all these data sources to arrive at a coherent and informative observation is the core of data fusion.

3.3.2 Autonomous navigation systems

Spacecrafts operate in extremely complex environments. With the increasing number of space missions, it is no longer possible for space missions to be dependent on the ground segment only. In critical phases of the mission, or when communication is cut off, these decisions must be made within seconds. Autonomous systems for navigation and planning functions for spacecraft are, therefore, increasingly implemented and integrated within space missions. These functions are driven by a combination of AI and ML systems that are capable of treating and processing the data and of making decisions.

Generally, in autonomous space systems, AI is necessary within the domains of perception and navigation; planning and learning.

In addition to navigation, spacecraft and space rovers could have a particular mission to accomplish, which might require a certain sequence of actions to be decided, for example performing repair and maintenance tasks. The spacecraft must be able to act deliberately to fulfil its mission. Maintenance or refuelling are good examples to illustrate the AI application for planning and learning. These tasks would involve navigating to the exact location of interest, observing the environment for collision-free handling of the maintenance/re-fuelling arm, and implementing the proper action for the desired outcome.

3.3.3 Spacecraft health monitoring

Ensuring the safety and reliability of a space system is one of the most critical concerns [33]. Anomalies and fault detection are crucial elements in ensuring the safe operation of a spacecraft in the harsh environment of space. In many cases, it is extremely difficult, if not impossible, to repair a spacecraft once it started its mission. Significant attention needs to be given to fault detection and diagnosis. Following the detection of a fault, the system must be able to trace the fault and isolate it.

Two clear applications of AI in anomaly and fault detection are

- Adaptive control limits for evaluating nominal operations of space systems, typically predefined before mission implementation.
- Ability to diagnose the system at a distance. Traditional methods rely on pre-programmed checks that should be run by a program to ensure the proper functioning of the system. However, these methods are incapable of detecting new and unseen faults or anomalies that might occur which have not been previously programmed by the operator. This is where AI can be useful.

3.4 Cyber vulnerabilities to AI in space

To summarise Section 3.3, Table 3 reports additional vulnerabilities deriving from the use of AI systems in space, in addition to the legacy vulnerabilities of Table 1. Referring to Fig. 1, a threat agent can recognize the AI system on-board as a threat target.

Table 3. Additional cyber-vulnerabilities from AI technologies in space

Source of vulnerability	Description of vulnerability
On-board remote sensing devices	Manipulation of databases during the learning phase leads to errant input recognition, suggesting a wrong output
System health check	System anomalies not detected or report of anomalies in a healthy system
Autonomous system	Improper use of mechanical arms causing damage to other space assets

It is to note that malicious AI algorithms can reduce performance and/or disrupt the nominal operation of benign AI algorithms as well [34]. A parallel work of the authors investigates the role of AI as threat agent [35].

In addition to the classical cyber-attacks, AI introduces the following typologies of threat actions (Section 2.1): data misclassification, synthetic data generation and data analysis [2].

3.4.1 Threat actions to AI in space

This section illustrates some examples of cyber-attacks deriving from the vulnerabilities of AI listed in Table 3.

For instance, let us consider an AI system which classifies images from satellites to detect specific buildings and suppose that the system is providing the operator critical information for the defence for detections of targets. The human operator would take consequent decisions based on the output of AI, without knowing the process behind such classification (black-box in Table 2). The operator might take decisions based on an erroneous output of AI potentially leading to catastrophic effects. The threat agent in this case is not easy to detect: was it due to the mathematical model of the AI failing to classify the building correctly because of, for example, environmental boundary conditions? Or was it because the AI system was subjected to data poisoning from an external agent? There is, therefore, a clear need to clarify and regulate the level of autonomy of the AI system in space, as well as its interaction with the human operators.

On-board fault management is also a critical point for AI applications striving for high autonomy, as associated with the vulnerability "System health check" of Table 3. An attack on spacecraft health monitoring systems may lead to a wrong identification of malfunctions in the system. This may lead to a variety of malfunctions going unnoticed, threatening the integrity of the objectives of the mission.

It is, therefore, of significant importance to properly address the new vulnerabilities that characterize the space systems and missions that are generated by virtue of AI. The following section outlines the current regulatory framework governing AI technologies applied in space, with the aim of making recommendations as to how to enhance cybersecurity for AI-driven space systems.

4. International governance of AI

The United Nations (UN) space treaty regime provides the basic rules and principles of space law that govern space activities [36]. The UN space treaties are drafted in a manner that allows to accommodate innovation of new technologies – this is because space activities are addressed in a general manner rather than in relation to any specific technology that might arise [37]. Hence, the broad principles under the UN space treaty regime would apply to the integration of new technologies into space assets [37], such as cybersecurity and AI. Broad legal principles though require legal interpretation. In this regard, soft law can help to interpret the law and determine what behaviour is compliant with the principles of law.

Albeit not legally binding [38], soft law, which covers technical standards among other non-binding practices and instruments [37], is part of the overall governance of space activities. Currently there is no technical standard that, as a regulatory tool, specifically deals with AI-driven space systems, let alone the cybersecurity for such systems. At international level, general AI governance is mainly achieved through soft law mechanisms, in particular standards and pre-standardisation documentations issued by the International Organisation for Standardisation and International Electrotechnical Commission (ISO/IEC), the Institute of Electrical and Electronics Engineers (IEEE), and the International Telecommunication Union (ITU), among others. Although cyber security of AI is subject to regulation within the respective standardisation frameworks, a comprehensive cybersecurity approach towards AI is still lacking.

4.1 Specificities of AI regulatory force

At regulatory level there is an acknowledgement regarding the importance which AI definition has when identifying its security issues. However, the technical regulatory framework does not provide for a uniform definition of AI – instead multiple definitions address different AI technologies. Such an approach may appear as a hindrance to standardisation of the entire AI ecosystem and contribute to the fragmentation of the applicable regulatory framework. However, regardless of the specific sector of AI application, the overall standardisation process is characterised by technical cooperation regarding the way standards are produced. Thus, despite being fragmented, AI regulation is developed through synergies between different sectors and organisations [46].

4.2 Recommendations regarding cyber security of AI space applications

A relevant regulatory framework through standardisation governing AI-driven space systems, which accommodates a comprehensive cybersecurity approach towards such systems, needs to emerge at international level. Such a soft law mechanism would provide a flexible form of international cooperation in the context of an increasing technological progress. Furthermore, it would enable a certain awareness of the underlying risk which the hard law may fail to do. Additionally, because it is less rigid than hard law, governance via standardisation would be sufficiently agile to adapt to innovation and ensure responsible behaviour when designing AI based applications.

When addressing cybersecurity of AI-based space applications, the above-mentioned regulatory framework needs to accommodate the non-deterministic nature of AI. The security of AI is directly associated with its level of uncertainty – hence the importance for the future cybersecurity regulatory framework governing AI-based space applications to aim for a certain degree of control over the uncertainties inherent to AI.

Another regulatory objective would be to incorporate technological rules which could be widely adopted throughout technical standards and guidelines. When doing so it is necessary to develop general principles addressing AI in the context of space autonomy as well as application-specific technical requirements or guidelines in parallel [39]. Thus, defining general requirements, such as integrity and availability, needs to be followed by detailed requirements that include concrete technical specifications assigned to AI-driven space systems, which connect cyber threats and vulnerabilities to said systems. When developing a cybersecurity framework specific to AI-driven space systems, potential future technological advancements need to be taken into account as new technological paradigms may emerge, such as high-level autonomy for space systems and subsystems *versus* human-robot interaction [40].

A regulatory framework governing cybersecurity of AI-driven space systems needs to be built upon common cybersecurity principles, such as the well-established

triad of CIA. This would be aligned with the current practices in the space industry where a minimum standard for information security is required to be followed.[1] Common cybersecurity principles also are included in security standards for space system engineering.[2] Building technology security upon them, although recommended from an engineering perspective, may have some drawbacks. In particular, the integration of such standards would reduce the overall vulnerabilities common for legacy space systems, such as those presented earlier in Section 2. Nonetheless, because they are not specifically related to AI, they are limited in addressing the new vulnerabilities generated by AI, i.e., as outlined in Section 3 of this article.

Consideration needs to be given to synergies between the space sector and other sectors where similar AI technologies are used [46] – such synergies are encouraged by the development of the current regulatory landscape as mentioned earlier in this paper. This would allow to use existing standards applicable to AI and IT environments but doing so against the specific threats and cyber vulnerabilities that are seen in AI space applications. Moreover, in order to enhance cybersecurity for AI-driven space systems a regulatory emphasis needs to be placed upon cyber resilience. Cyber resilience is built upon known vulnerabilities of AI-driven space systems and provides for some flexibility for future vulnerabilities. It is also important to consider the level of criticality of the specific component or subsystem where AI is applied.

4.2.1 Preventive mechanisms

The regulatory framework concerning cybersecurity of AI-driven space systems needs to incorporate a prudent balance between *a controlled degree* of uncertainty inherent to AI, on one hand, and, on the other hand, advances in autonomy, optimisation and robustness capabilities achieved through AI [14]. In relation to this, account should be taken of the specific needs for such capabilities resulting from the operation of complex space systems [14]. A trade-off could be achieved at the design stage by implementing a layered security architecture of the AI-driven space system, which may accommodate some level of uncertainty related to AI [41]. This can be done by ensuring that, when designing the AI model, alternative security mechanisms apply at the upper level if the AI system signals too much uncertainty at the lower level [41]. This mechanism has been developed in relation to AI applications in the automotive industry and, as such, could serve as a starting point to develop alternative security mechanisms for AI space applications. In this regard, generally applicable methods for determining the uncertainty of models are needed.

Another mechanism which would allow a controlled degree of uncertainty is testability by design. Verification and validation, as well as testing and evaluation methods, play a central role. Validation and verification take place during the development stage of an AI space system and provide formal checks of the AI module and its application [42]. Regardless of the specific technique, testing of an AI system on a large number of combinations of environmental conditions facilitates a characterisation of the predictable behaviour of said system – thus decreasing the inherent uncertainty of the system [41].

As part of the design of the AI space application, the testability process includes quality requirements. In the standardisation and pre-standardisation literature there is an agreement that quality requirements regarding AI systems are directly related to security of said systems [41, 42, 43]. In connection with this, quality assessment contributes to reaching a controlled degree of uncertainty inherent to AI, namely by scrutinising the quality of data. Considering the specific data driven nature of AI space systems, data have a significant impact upon the behaviour of such systems. The current general data quality framework applicable to data processed in a structured way with a computer system, namely the ISO/IEC 25012 is not entirely relevant to AI-driven space systems. However, it provides for an assessment model, which could serve as guideline regarding quality assessment specific for AI space applications.

Based on the above, testability by design enables cyber security by design. In light of this paradigm, the validation and verification process as well as the quality requirements constitute preventive measures for the purposes of enhancing cyber resilience of AI space applications.

Other preventive measures include risk management and ongoing evaluation of the cyber security posture of AI-driven space systems. Risk management, as a regulatory tool, is designed to address uncertainties[3] – thus it could help to reach the balance between a tolerable level of risk and the non-deterministic behaviour typical for AI systems.

4.2.2 Mitigation measures

The role of security engineering consists in reducing the probability of a threat to occur or reducing its consequences on the system and mission level. Typical technology solutions implemented as mitigation against cybersecurity threats include, but are not limited to, encryption for TT&C, and zero trust architecture [44].

[1] Such as CIS Critical Security Controls® v8, 2021.

[2] Such as ISO 20214:2015 Space data and information transfer systems – Security architecture for space data systems.

[3] According to ISO 31000:2018 Risk management - Guidelines, risk is defined as "effect of uncertainty on objectives"

These technological solutions, albeit not AI specific, are effective methods to increase the difficulty for the adversary to conduct a successful cyberattack on the system.

Despite current technological solutions to protect space assets from threat agents, no space mission is completely safe from cyber threats. It is crucial that mitigation and recovery processes are available in case of a cyberattack. Implementing appropriate mitigation measures regarding cyberattacks targeting AI is conditional upon preliminary classification of the AI threats [43]. In any case, the focus of the mitigation and recovery processes needs to be on enabling speed recovery process which includes characterising the threat classes immediately, releasing mitigations into the network to counter cyberattacks in an emergency change control process, and engaging a human supervisor [45].

Finally, mitigation processes need to be adapted over time since new conceivable attack vectors require corresponding defence and mitigation mechanisms to be constructed.

5. Conclusions

The use of AI in space applications is a trend predicted to increase. This is due to the growing interest in process optimisation and automated services, as well as in reducing the elaboration time for data and imaging processing in the field of space industry. However, AI exhibits cyber vulnerabilities and, thus, is not exempt from cyber threats.

The first part of this research shows how AI technologies for space applications can potentially increase the cyber-vulnerability of the system because of new threat targets associated with the technology itself. While it is clear from the literature review that the ground segment and the command paths are the weakest access points in legacy space systems, AI provides additional possibilities for a threat agent to manipulate the data at different stages of its lifecycle. In fact, due to the complex interaction between hardware and software architectures, an external observer would interpret the AI-system as a "black-box" and rely on its output with a high degree of confidence. Database manipulation on remote sensing devices, undetected system anomalies during systems health checks or improper use of mechanical appendices in autonomous systems are only examples of events which could threaten AI-systems in space.

Considering the rapid technological advancements in the space industry, the lack of cybersecurity protocols and regulations become an important topic for the use of emerging technologies in space. AI-based space applications are under development, and as with many other technologies to be used in space, the cybersecurity of these assets is a priority for the safety of the mission data and the security of future missions. Therefore, the common ground for regulations will certainly have technological and legal dimensions. One of the focal points of a regulatory framework specific for AI-driven space systems is suggesting preventive and mitigation measures, as a way to establish a certain degree of control over the uncertainties inherent to AI. To do so, it is necessary to contextualise AI in the field of autonomous systems in a general manner, including non-space industries of AI applications. The preventive approach shall ensure the development of cyber resilient systems by design, while mitigation measures would diminish the consequences of occurrence of a threat on the system. Lastly, developing cybersecurity principles which are adaptable to new and emerging technologies and their use in space, by considering both mission-specific environments and the new threats, is as important as developing and monitoring best practices.

Acknowledgements

The authors gratefully acknowledge the expertise and valuable support of the Space Generation Advisory Council (SGAC) - "Space and Cyber Security Project Group".

References

[1] Report Concerning Space Data System Standards - Security Threats Against Space Missions, The Consultative Committee for Space Data Systems, CCSDS 350.1-G-3, February 2022.

[2] The future of the European space sector - How to leverage Europe's technological leadership and boost investments for space ventures, European Investment Bank, 2019.

[3] Guide for Conducting Risk Assessments, NIST Special Publication 800-30, Revision 1, September 2012.

[4] INCOSE - International Council for Systems Engineering, Systems Security Engineering: Mission and Objectives, July 2021.

[5] Open Threat Taxonomy Definition, J. Tarala and K. Tarala (Eds.), v1.1, Enclave Security, 2015. https://www.auditscripts.com/resources/open_threat_taxonomy_v1.1a.pdf (accessed 31.08.2022).

[6] Report Concerning Space Data System Standards - Security Threats Against Space Missions, The Consultative Committee for Space Data Systems, CCSDS 350.1-G-3, February 2022, https://public.ccsds.org/Pubs/350x1g3.pdf (accessed 31.08.2022).

[7] NSR, Cybersecurity Report, April 2022, https://www.nsr.com/nsr-space-cybersecurity-white-paper/ (accessed 31.08.2022).

[8] Ground Data Systems and Mission Operations, NASA, 2021.https://www.nasa.gov/smallsat-institute/sst-soa/ground-data-systems-and-mission-operations (accessed 31.08.2022).

[9] National Institute of Standards and Technology, Security Requirements for Cryptographic Modules, Federal Information Processing Standard (FIPS), ISO/IEC 19790:2012; ISO/IEC 24759, 2014.

[10] Space System Protection Requirements, NASA STD-1006 w/Change 1, 2019. https://standards.nasa.gov/standard/NASA/NASA-STD-1006 (accessed 31.08.2022).

[11] Hamon, R., Junklewitz, H. and Sanchez, I., Robustness and Explainability of Artificial Intelligence, Publications Office of the European Union, 2020.
[12] Russell, S. J., Artificial intelligence a modern approach. Pearson Education, Inc., 2010.
[13] Izzo, D., Martens, M., Pan, B,. A survey on Artificial Intelligence Trends in Spacecraft Guidance Dynamics and Control, in Astrodynamics, *3*(4), 2019, 287-299.
[14] Jaekel, S., B. Scholz, Utilizing Artificial Intelligence to Achieve a Robust Architecture for Future Robotic Spacecraft, 2015 IEEE Aerospace Conference, 07-14 March 2015.
[15] Bertino, E, Attacks on Artificial intelligence [Last Word], in IEEE Security & Privacy, Volume 19, Issue 1, Jan-Feb 2021, 103 - 104.
[16] Beed, J.M., Rogers, W.A, Fisk, A,. Towards a Framework for Levels of Robot Autonomy in Human-Robot Interaction, in Journal of Human-Robot Interaction, 2014, pp. 74 - 99.
[17] Girimonte, D. and Izzo, D., Artificial intelligence for space applications, in Intelligent Computing Everywhere, Springer, London, 2007, 235-253.
[18] Manning, J., Langerman, D., Ramesh, B., Gretok, E. Wilson, C.M, George, A.D., Mackinnon, J. and Crum, G., Machine-Learning Space Applications on SmallSat Platforms with TensorFlow, 32nd Annual AIAA/USU Conference on Small Satellites, 04-09 August 2018.
[19] Rai, L., Kook, J., Hong, J., Non-Deterministic Behaviour Modelling Framework for Embedded Real-Time Systems Operating in Uncertain Environments, in Journal of Information Science and Engineering, Vol. 26(1), 2010, 83-96.
[20] Tyukin, I.Y., Higham, D.J., Gorban, A.N., On adversarial examples and Stealth Attacks in Artificial Intelligence Systems, *in* 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, United Kingdom, 2020.
[21] Vorm, E.S., Assessing demand for transparency in intelligent systems using machine learning, *in* 2018 Innovations in Intelligent Systems and Applications (INISTA), IEEE, July 2018, 1-7.
[22] Lockey, S., Gillespie, N., Holm, D. and Someh, I.A., A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities and Future Directions, Hawaii International Conference on System Sciences, 2021.
[23] Huang, S., Papernot, N., Goodfellow, I., Duan, Y. and Abbeel, P., Adversarial Attacks on Neural Network Policies, 2017.
[24] Qiu, S., Liu, Q., Zhou, S. and Wu, C., Review of artificial intelligence adversarial attack and defense technologies, in Applied Sciences, 9(5), 2019, .909.
[25] Lorenzo, P,. Stefano, F., Ferreira, A., and Carolina, P,. Artificial Intelligence and Cybersecurity: Technology, Governance and Policy Challenges, Centre for European Policy Studies (CEPS), 2021.
[26] Rajan, K. and Saffiotti, A., Towards a science of integrated AI and Robotics, in Artificial Intelligence, Vol. 247, 2017,1-9.
[27] Oche, P.A., Ewa, G.A. and Ibekwe, N., Applications and Challenges of Artificial Intelligence in Space Missions, in IEEE Access, 2021
[28] Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G. and Johnson, B.A., Deep learning in remote sensing applications: A meta-analysis and review, ISPRS Journal of Photogrammetry and Remote Sensing, Vol. *152*, 2019,166-177.
[29] Furano, G., Meoni, G., Dunne, A., Moloney, D., Ferlet-Cavrois, V., Tavoularis, A., Byrne, J., Buckley, L., Psarakis, M., Voss, K.O. and Fanucci, L., Towards the use of artificial intelligence on the edge in space systems: Challenges and opportunities, *in* IEEE Aerospace and Electronic Systems Magazine, Vol. *35*(12), 2020, 44-56.
[30] Girimonte, D. and Izzo, D., Artificial intelligence for space applications, *in* Intelligent Computing Everywhere, Springer, London, 2007, pp. 235-253.
[31] Meß, J.G., Dannemann, F. and Greif, F., Techniques of artificial intelligence for space applications-a survey, European workshop on on-board data processing (OBDP2019), European Space Agency, February, 2019.
[32] Jeppesen, J.H., Jacobsen, R.H., Inceoglu, F. and Toftegaard, T.S., A cloud detection algorithm for satellite imagery based on deep learning., *i*Remote Sensing of Environment, Vol. 229, 2019, 247-259.
[33] Ingrand, F. and Ghallab, M., Deliberation for autonomous robots: A survey, in Artificial Intelligence, 247, 2017, 10-44.
[34] Yamin, M. M., Ullah, M., Ullah, H., Katt, B., Weaponized AI for cyber attacks, in Journal of Information Security and Applications, Vol. 57, March 202
[35] Mantı, N.P., Carlo, A., Markova, R., Jha, D., Breda, P., Abdin, A., AI systems to ensure cyber security in space, International Astronautic Congress (IAC), 2022.
[36] Masson-Zwaan,T., Hofmann, M., Introduction to Space Law, Wolters Kluwer, Fourth Ed., 2019, 40.
[37] Blount, P.J., A Satellite is Just a Thing on the Internet of Things, in Air and Space Law, Vol. 42, 2017, 273-293.
[38] Ferrazzani, M., Soft Law in Space Activities – An updated view, in Irmgard Marboe (Eds), Soft Law in Outer Space - The Function of Non-binding Norms in International Space Law, Böhlau Verlag, Köln/ Wien, 2012, 99-117.
[39] Kunze, L., Hawes, K., Duckett, T., Hanheide, M., Krajnik, T., Artificial Intelligence for Long-Term Robot Autonomy: A Survey, in IEEE Robotics and Automation letters, Vol. 3, No 4, October 2018, 4023 - 4030.
[40] Gao, Y., Introduction, in Yang Gao (Eds.), Space Robotics and Autonomous Systems Technologies, Advances and Applications, The Institution of Engineering and Technology, 2021, 1 - 10.
[41] DKE German Commission for Electrical, Electronic & Information Technologies of DIN and VDE, German Standardization Roadmap on Artificial Intelligence, November 2020, 75 - 111.
[42] ISO/IEC TR 24028, "Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence", First edition, 2020-05, 6 - 31.
[43] ETSI, Securing Artificial Intelligence (SAI); Problem Statement, ETSI GR SAI 004 v1.1.1, 2020-12, pp. 10-11.
[44] IBM, Zero Trust Architecture, https://www.ibm.com/topics/zero-trust (accessed 31.08.2022).
[45] Vandermeulen, R., Space – Cyber Governance Project Workshop, in The Vincent and Elinor Ostrom Workshop, June14, 2022, https://www.youtube.com/watch?v=cuUiFDN-tX0 (accessed 31.08.2022).
[46] Markova, R., AI in the Context of Space Autonomy, SGAC Space and Cyber Security Project Group, Space and Cyber Brief Series, February 2022, https://spacegeneration.org/wp-content/uploads/2022/04/SC-Brief-Series-2-AI-1-2.pdf (accessed 31.08.2022).