SpaceOps-2025, ID #49

Innovating Space Operations with AI: The AISHGO Project

Nicola Policella^{a*}, Jose Martinez^a, Nieves Salor Moral^b, Sakis Kotisis^b, Kathrin Helmsauer^c, Tobias Göttfert^c, Marco Cristoforetti^d, Andrea Gobbi^d, Florian Hegwein^e, Gabriele De Canio^{f*}, Age-Raymond Riise^f, Sinda Mejri^f, Viet Duc Tran^f, Ignacio Tanco^f

- ^a Solenix Engineering GmbH, Germany, {nicola.policella} {jose.martinez} @solenix.de
- ^b Starion Group, Germany, n.salor@stariongroup.eu
- German Space Operations Centre, DLR, Germany, {kathrin.helmsauer} {tobias.goettfert} @dlr.de
- ^d Fondazione Bruno Kessler, Italy, {agobbi}{marco.cristoforetti}@fbk.eu
- e Airbus, Germany, florian.f.hegwein@airbus.com
- f European Space Operations Centre, European Space Agency, Germany, name.surname@esa.int
- * Corresponding Author

Abstract

Artificial Intelligence (AI) is increasingly permeating and improving various aspects of our daily lives; but can it also revolutionize how spacecraft are operated? This is the central question driving the project "AI for Automation of Satellite Health Monitoring and Ground Operations (AISHGO)." The project aims to integrate AI into mission control environments, such as the European Space Agency's European Space Operations Centre (ESOC) DLR's GSOC, to automate and optimize satellite health monitoring and ground operations.

AISHGO is structured around four pivotal use cases, each addressing distinct aspects of satellite operations. These include machine learning-based incident classification and root-cause analysis assistance, AI-based predictive maintenance, intelligent telemetry data anomaly detection, and AI-based long-term satellite health monitoring. The research leverages both structured and unstructured data to develop data-driven AI solutions that surpass classical models in efficiency and effectiveness. By utilizing advanced AI techniques and models such as transformers, large language models (LLM), and long-short-term memory (LSTM) networks, the project demonstrates the capability of AI to predict system behaviors, flag potential issues, and enable operators to take preventative measures.

The solutions proposed have been validated through both simulations and real-world scenarios, confirming their practical value and highlighting the feasibility of adopting AI technologies in modern space operations.

Keywords: Artificial Intelligence, Spacecraft Operations, LLM, Predictive Maintenance, Automation

Acronyms/Abbreviations

Artificial Intelligence (AI)

AI Powered Incident Assistant solution (AIIA)

AI for Automation of Satellite Health Monitoring and Ground Operations (AISHGO)

European Space Agency (ESA)

European Space Operations Centre (ESOC)

Fault Detection, Isolation, and Recovery (FDIR)

Graphical User Interface (GUI)

German Space Operations Centre (GSOC)

Key Performance Indicator (KPI)

Machine Learning (ML)

Minimum Viable Product MVP)

Out Of Limit (OOL)

Spacecraft (S/C)

Satellite health Monitoring AI-powered suite (SMAI)

Telemetry (TM)

1. Introduction

Artificial Intelligence (AI) has demonstrated its value in the automation of processes across a wide range of domains, from finance and healthcare to the automotive industry, and more recently, space operations. In 2021, the European Space Agency, in collaboration with industry partners, launched the *Artificial Intelligence for Automation* (A^2I) Roadmap [1,2]. This initiative identified a comprehensive set of use cases across all phases of mission operations,

Published by the Canadian Space Agency (CSA) on behalf of SpaceOps, with permission and released to the CSA to publish in all forms."

each evaluated and prioritized according to its expected impact and technical readiness. The roadmap is guiding the development of AI applications targeting key operational areas such as predictive maintenance of ground systems, automated mission planning, simulation and validation of operational scenarios, spacecraft health assessment, anomaly detection and diagnosis, and decision-support tools for flight controllers [3-9].

Within this framework, the AISHGO project focuses on validating AI across four distinct use cases: incident investigation and root-cause analysis, predictive maintenance, data anomaly detection, and long-term satellite health forecasting [10]. The primary objective of AISHGO is to develop and implement AI-driven tools that assist satellite operators at the European Space Operations Centre (ESOC) and other satellite operators. These tools are designed to be user-friendly, versatile, and applicable to a broad range of satellite missions.

Throughout the project, we identified not only common needs across all the use cases but also shared elements among domain-specific use cases. This overlap enabled us to consolidate various functionalities into a smaller set of operational tools. For instance, users can conduct anomaly detection and long-term analysis with a single solution, eliminating the need to switch between tools. Additionally, users can predict maintenance tasks for ground station subsystems and continue their investigation using intelligent assistants within the same application. Furthermore, our analysis explored users' expectations in detail and assessed how AI integration could reshape the operational landscape, including potential workflow modifications and the acceleration of decision-making processes. We quantified the impact of these tools using Key Performance Indicators (KPIs).

Section 2 provides an overview of these four use cases, while Section 3 explores the applications and prototypes developed for each use case, starting with an analysis of their crucial aspects. Section 4 summarizes the AI approaches investigated to address these challenges. Finally, Section 5 discusses the two resulting tools, AIIA and SMAI, before concluding the paper.

2. Use cases

The AISHGO project focused on validating AI across four different use cases, ranging from incident investigation and root-cause analysis to predictive maintenance, data anomaly detection, and long-term satellite health forecasting. The following subsections provide a concise description of the different use cases.

2.1 Machine learning based incident classification and root-cause analysis assistance

Ground station and subsystem engineers investigate incidents based on daily incident logs received via email. However, because investigation timing depends on work shifts and geolocation of the ground station across the globe, critical incidents may face unnecessary delays. Automating incident classification, providing an intelligent initial contextual enhanced investigation and implementing an alert system would prioritize high-risk issues, reducing response time.

Currently, incident processing and internal communication rely heavily on emails, scattered siloed tools (e.g., ARTS, OSPMS, documents) and chats between the engineers. Thus, making it difficult to track recurring issues and lessons learned. This inefficiency extends to the weekly review meetings, where critical, unresolvable or difficult incidents are often re-investigated manually by more expert personnel, leading to redundant effort. An AI engineer would streamline the investigation and review processes by providing initial assessment of the incident abased on predefined criteria (e.g., investigation duration, impact, subsystem involvement, problem consequences, resolution), automatic classification and offering a list of similar problems, while creating a knowledge base over a structured system storing investigations and engineers' interactions.

While the number of missions are growing almost exponentially each year, the ground stations are not scaling at the same pace. Thus, incidents become more frequent and harder to manage. The current manual approach will soon be unsustainable, necessitating a more intelligent, automated and structured investigation process.

2.2 AI-based predictive maintenance

ESA aims to leverage AI to analyze time-series data and logs to predict maintenance needs before issues arise in different ground station subsystems. Given their complexity, the pilot project focuses on the Frequency & Timing subsystem (F&T), which supports other ground station systems. Specifically, the pilot targets the Deep Space ground stations using Active Hydrogen Maser (AHM) atomic clocks for time synchronization.

The F&T system ensures precise frequency (10 MHz) and timing (1 PPS) for transmitters, receivers, and synchronization between ground stations. At its core, the AHM atomic clock, steered to UTC, maintains accuracy but still experiences slight drifts due to factors like temperature fluctuations and resonant frequency changes. While housed in a temperature-controlled environment (+/- 0.5°C), periodic corrections are needed. However, monitoring of these small drifts is a repetitive task still being done manually against the system specifications by a reduced number of expert engineers in combination with the assessment of annual reports generated by the hardware providers on how

Published by the Canadian Space Agency (CSA) on behalf of SpaceOps, with permission and released to the CSA to publish in all forms."

improvements need to be made after problems have already taken place. Furthermore in 2024 a replacement of the F&T system was rolled out in different stations making automatic data analysis difficult to extrapolate results and leverage the existing procedures.

By implementing an AI-driven predictive maintenance and reporting system, ESA aims to enhance system reliability and minimize downtime of each subsystem to later scale it to the rest to also analyze impacted problems across the whole ground station and ground control system

2.3 Intelligent telemetry data anomaly detection

Spacecraft operators need to monitor the performance of the spacecraft and make sure that it can continue being productive (e.g. delivering science data). An important aspect of the performance monitoring activities consists of verifying the expected behavior, detecting and dealing with known anomalies and detecting potential anomalies.

A common approach to do anomaly detection on-ground is to implement Out-Of-Limits (OOL) checks. There can be various OOLs for a single TM parameter. In the simpler case, it corresponds to "soft" and "hard" limits, depending on how far they go, typically the hard limit being wider than the soft limit. To bring some context, some missions can also define conditional OOL thresholds. The condition corresponds to the raw value (or range of raw values) of another TM parameter. There are several applications for the conditional OOL; a typical example consists of TM parameters behaving differently depending on the working mode of a unit. The Mission Control System performs these checks and alerts operators when any OOL triggers. There are also machine-learning based approaches to help operators in performing anomaly detection. When a first-time anomaly occurs, operators need to investigate it and devise a plan to deal with it.

The first-time anomalies are more the exception than the norm. This use case focuses on helping operators with their daily tasks of verifying nominal operations and detecting and reporting known anomalies. The known anomalies have already been assessed by the flight control team (FCT) and they know how to properly identify them. They need help in monitoring their occurrence, receiving notifications and creating reports.

2.4 AI-based long-term satellite health forecasting

To ensure spacecraft safety, operators must estimate future physical quantities, such as the expected power consumption of the thermal subsystem. While some physical models, such as degradation models for batteries and solar arrays, are available to the flight control team, they do not fully meet all mission needs.

The goal of the use case is to enable long-term monitoring (multi-year) by leveraging historical data and conditions (e.g., eclipses, variations in the spacecraft's distance from the Sun, etc.). The long-term analysis may notice degradations in comparable situations ahead of time that would need to be brought to the attention of the FCT. Detecting broken trends is also relevant in monitoring long-term performance.

In addition, long-term forecasts can support long-term mission planning and, by comparing predictions with actual data as it becomes available, allow operators to identify significant deviations that may require investigation. Additionally, analyzing discrepancies between long-term predictions and actual data will help assess and refine the accuracy of forecasting models, ultimately improving spacecraft health monitoring and operational decision-making.

3. Prototyping AI Solutions

During the initial phase of the AISHGO project, special focus was given to demonstrate that AI can be strategically leveraged to support and achieve the goals of various use cases. By tailoring AI techniques to address the specific needs of each scenario, we made sure that the proposed solutions were both effective and contextually relevant. The initial scoping and inception phase was not only focused on exploring AI's potential benefits but also on proactively mitigating risks associated with its deployment.

One of the core challenges in AI adoption is managing the risks inherent to its use. To address this, we developed, where possible, explainable and trustworthy AI solutions (see details in Sect. 4). Ensuring that AI-driven insights could be interpreted by end-users was a key objective. This approach supports a responsible adoption of the AI solution while minimizing possible concerns.

Throughout the process, user feedback played a pivotal role in shaping and refining the AI solutions. Engaging with end-users enabled us to iteratively improve the feature sets associated with each Minimum Viable Product (MVP). This iterative validation process ensured that the MVPs were well-aligned with user expectations and functional requirements, ultimately paving the way for the final product development.

3.1 Analysis of the use cases objectives, common KPIs, specific aspects

In this section each use case is analyzed with a focus on the specific objectives. KPIs have also been defined to assess the quality of both the initial Minimum Viable Products (MVPs) and the final solutions for each use case.

3.1.1 A1 - Machine learning based incident classification and root-cause analysis assistance

The objective of the A1 use case is to build a knowledge and investigation tool revolving around the concept of a "knowledgeable and reasoning" artificial intelligent ground station engineers to help on the investigation of incidents. The tool should offer precise and reasoned answers while identifying the used sources to build the trust while reducing the hallucination innate to Generative AI.

The MVP of A1 revolves around the idea of an incident investigation. For each incident an investigation exists. As such, the MVP required to structure the investigations and record the station engineer workflows so they could later be revisited and review at any moment in time snapshotting the work. Upon occurrence of the incident, it is classified under a category of the possible root-cause by the engineer that has detected the problem. This classification can change afterwards by more experience engineers or after investigation has been raised more information. However, there is no official record on why the category has changed or whether incident has occurred in the past or if similar ones have happened (either in the same ground station or in others). Thus, each investigation had to record the answers to these typical questions each engineer looks for in logs, report, specifications, mails, etc. while automating the classification and finding similar incidents. So, although core, the A1 MVP could not be only a fine-tuned LLM answering questions, but an agentic ecosystem of both AI and software tools that collaborate to bring up knowledge in each investigation.

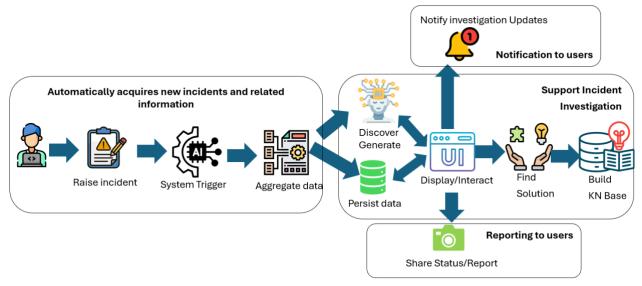


Figure 1: A1 workflow.

But, as with any AI related solution, one of the most difficult aspects of the solution was that it should support big amounts of data both in archive (as historical incidents and solutions) and streaming modes working for all ground stations across the globe, handling different levels of expertise with subjective sources (defined in natural language by the engineers themselves), but also automatic timestamped logs from the stations subsystems.

During the MVP we focused on de-risking the AI agents to demonstrate similarity vector searches and clustering approaches, correct classification and the LLM to achieve answers that were well reasoned, pointing to real and meaningful sources and adequate. During this phase experimentation with different open source LLM models, parametrization sizes, vector databases, and AI agent pipelines were tested. Finally, the MVP only ingested archive data of past incidents of a couple of years while generating an End-to-End Prototype providing all previously defined capabilities in a reduced infrastructure with low GPU capabilities.

To evaluate the impact of the solution, the following Key Performance Indicators (KPIs) were considered:

KPI	Description
User Adoption	Degree to which the tool can be adopted by a wide number of users, and they keep using the tool
	once tested

Reduce of latency	Degree to which the investigations can be automatically initiated as soon as the problem is detected, and initial steps and questions are available when engineers start their shifts to reduce the time spent on performing typical discovery processes.
Indication of trends on the issues	Degree to which the past incidents can be identified as similar to learn actions taken from those and possible causes.
AI support	Degree to which the tool can reason about the failure modes, retrieve the relevant data and present it in a structured manner and support the conclusion on the cause of the error.
Contextual Information	Reduce on the number of data files (and thus time) the engineers have to check for tracing relevant information.
User Experience	Intuitive and user-friendly interface with an easy learning curve.

3.1.2 A2 - AI based predictive Maintenance

A2 Use Case main objective is to build a flexible but intuitive web-based tool for predictive maintenance which can be applied to the different subsystems of the ground stations and could be later integrated into the rest of the ground stations software (e.g. GSM-CC for ESOC). The solution should be able to reduce as much as possible both false positives (detections which are not real) and false negatives (lacking a detection for a real subsystem problem).

The MVP of A2 provides ground station engineers a side optional mechanism with the functionality to see predictions on the F&T subsystem potential problems based on the recorded monitoring data. It also allows users to visualize the detected potential problems, the raw data used for generating that detection and request and/or visualize any report with the contained information. Users can also interact with the solution configuring the detection they are interested into and further investigate the detection and perform maintenance actions recorded.

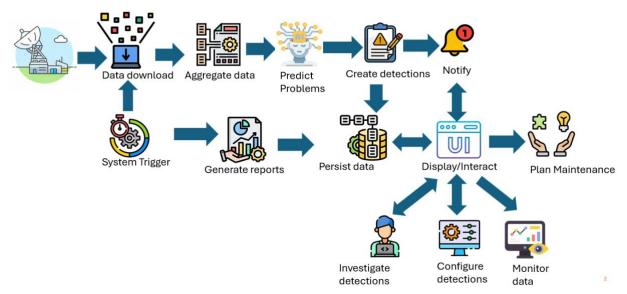


Figure 2: A2 Workflow.

The MVP solution is centered around the concept of detections and their configuration defined by station engineers knowledgeable of each subsystem and their specifications. For this, the work started by experimenting with data from Cebreros and Marlague F&T systems of the last 5 years (2019 to 2024) and using the known problems that have occurred based on the provider's reports of the maser clocks and the maintenance tickets opened on time. Although several AI approaches such as pattern matching or autoencoders were assessed (a full summary can be found on section 4), at the end and due to not having enough nominal data (i.e., the data not mostly did not match the subsystem nominal specifications or behave anomaly constantly), it was required to split the solution in 3 types of detections: predictive detections for the PPS (Precise Positioning Service) values which range between +200 and -200, flag trend disruptions over allowed range limits (and thus, highlighting a worsening of the quality of the subsystem) and monitoring the raw telemetry data over specification limits. Using these types of detections during the MVP phase, already detected several

problems that indeed happened in the past and even one of them that the F&T did not realise at the time with an average of 1 week time of advance and that in 2024 it led to the F&T to roll out a new generation of their masers starting in Cebreros.

On top of the AI detections, the ground station F&T managers also required custom reporting and notifications of the detections to be able to plan the maintenance and monitor the mitigation actions and their effects by recording actions taken and then continue monitoring the data for each detection to see whether the actions indeed solved the problems or how much time had been required to later calculate costs.

To evaluate the impact of the A2 MVP, the following Key Performance Indicators (KPIs) were considered:

KPI	Description
Reduction of the	Degree to which the station subsystems can reduce their downtime because the potential
overall ground	problems are detected before they happen and can be solved before.
station	
subsystem	
downtime	
Predictive	Measure the detection generated on a time period are correctly predicting a problem that could
accuracy	have occurred by measuring it in shadow mode until trust is achieved.
Time reduction	Improve the time from the creation of a potential issue detection including the issue expected
of the issue	occurrence time and performing all relevant workflow actions (i.e. investigation start, action
management	planned, repair action performed, confirmation of the resolution, close-up investigation).
workflow	
Number of users	Number of people defining detection type signatures and detecting potential problems.
Scalability	Can the solution be applied to other ground station subsystems?
User Experience	Intuitive and user-friendly interface with an easy learning curve.

3.1.3 F1 - Intelligent telemetry data anomaly detection

The objective of the F1 use case is to build a consolidated, user-friendly tool for Intelligent Telemetry Monitoring that is available to many flying missions. The tool should overcome the limitations of the current OOL approach discussed in section 2.3.

The MVP of F1 revolves around the concept of signatures. Signatures consist of a combination of indicators that specify the conditions that operators need to recognize, either nominal behaviour or recurrent anomalies. The indicators available during the MVP were event packets, telecommands and telemetry. Either occurrence or non-occurrence. For instance, a signature could be defined as the occurrence of an event and the non-occurrence of a telecommand in a relative time with respect to the event; this way, the reconfiguration of an on-board element could be detected without the FCT having commanded it.

Once operators define signatures, they go to the signature library. In the signature library, users can enable monitoring, notifications or make a signature public so that it can be seen by their colleagues. A signature can be run on past data to check previous occurrences or, when set to be monitored, it is scheduled to run in the future when data is downlinked. The notifications are available via email and on the tool itself. The dashboard shows a summary of the matches with statistics.

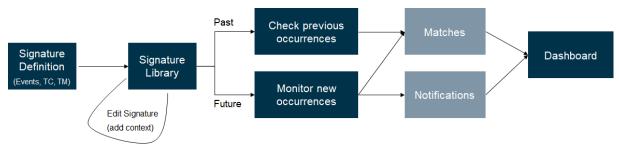


Figure 3: F1 Workflow.

During the MVP we focused on getting a first prototype to validate that we got the concept right and that we were developing the functionality that will be truly useful not only for the Product Owners, but also for most of the missions. We deliberately left out of the MVP the functionality that we knew with high degree of confidence that it will work

Published by the Canadian Space Agency (CSA) on behalf of SpaceOps, with permission and released to the CSA to publish in all forms."

such as email notifications, user authentication, etc. The MVP focused on validating the usefulness of use case specific functionality. The MVP was developed as a stand-alone application.

During the MVP phase, we identified potential usages for Artificial Intelligence: pattern matching and context suggestion. Pattern Matching in telemetry parameters is relevant to increase the expressiveness of the signatures, allowing operators to specify behaviours in combination with TM parameters. The Context Suggestion AI functionality allows operators to receive suggestions on which other TM parameters might be involved in the phenomena described by signatures. Due to the short duration of the MVP phase, the AI functionalities remained conceptual until the operationalization phase.

To evaluate the impact of the F1 MVP, the following Key Performance Indicators (KPIs) were considered:

KPI	Description
Scalability	Degree to which the tool can be adopted by a wide number of users.
Flexibility	Degree to which the tool can be used for a wider set of scenarios
Immediacy	How much increased automation can the solution introduce instantaneously compared to without
	it today?
Time saved	What Anomaly/TM check related tasks currently consume how much time and how can this tool
	help in reducing it. This also includes reporting.
AI integration	How much potential do the users see in adding assisting AI to the basic user assisting functions
	the tool envisages to provide?
Improve follow-	Improvement to operator's and manager's daily and follow-up tasks in routine and non-routine.
up tasks	
User Experience	Intuitive and user-friendly interface with an easy learning curve.

3.1.4 F3 - AI-based long-term satellite health forecasting

The F3 use case aims to reduce operator workload by automating Long Term Monitoring (LTM) analysis and reporting. It identifies stable trends and deviations automatically, minimizing manual tasks and providing intuitive tools to help operators monitor critical telemetry data throughout long-term missions.

The MVP provides the means to automatically analyze and detect changes in the trends of TM parameters. Detected changes are visualized to the user on a dashboard along with a feature rich graphical plotting tool for further investigation. AI is used to automatically find stable trends and their changes.

To evaluate the impact of the F3 MVP, the following Key Performance Indicators (KPIs) were considered:

KPI	Description
Time saved	Factor to measure by much the duration of the Trend Detection and Long-Term Report generation can be reduced.
Flexibility	Can a report/trend detection which is generated on demand, automatically and more often yield better results?
Quality	Improved report/trend detection consistency due to automatic report generation / trend detection
Number of users	Number of people producing LTM analysis and consuming LTM insights and reports
Scalability	Can the approach be transferred to other missions?
User Experience	Intuitive and user-friendly interface with an easy learning curve.

3.2 Moving Toward Operationalization

As the project transitions from ideation to operationalization, several key recommendations have been identified to enhance efficiency in achieving project goals. These recommendations encompass best practices for AI deployment, strategies for scaling solutions, and approaches to continuous monitoring and improvement. Additionally, for each use case, we identified specific areas of technical debt that require further attention. These identified gaps served as an initial backlog for the next phase of development, ensuring that critical issues are systematically addressed before full-scale implementation.

4. AI Approaches

In the context of aerospace ground operations, artificial intelligence offers a transformative potential for both retrospective incident investigation and proactive maintenance planning. A comprehensive framework has been developed that integrates Large Language Models (LLMs) with machine learning-based anomaly detection and

forecasting techniques. The system is designed to support engineering teams in understanding complex incidents, retrieving relevant historical information, and predicting potential system degradation before failures occur.

The framework comprises two primary components: an LLM-based assistant for incident analysis and a suite of predictive maintenance tools grounded in time-series analysis. The LLM assistant, based on LLAMA3 8B [11], is built on the Retrieval-Augmented Generation (RAG) paradigm, enabling contextualized, document-grounded responses to user queries. A knowledge base is constructed from heterogeneous textual sources such as incident logs, anomaly reports, ground station journals, emails, and chat records to achieve this. These documents undergo a structured embedding process, where they are first cleaned, segmented into meaningful text chunks, and transformed into high-dimensional vector representations that are saved into a dedicated Weaviate vector database [12]. These vectors preserve the semantic structure of the text and enable similarity-based retrieval of contextually relevant information.

The assistant architecture is composed of multiple stages. Initially, the system determines the most appropriate collection to query based on the user's intent. A filtering mechanism is then applied to refine the candidate set of documents using feature-level metadata. Retrieved documents are evaluated for their relevance to the query before being passed to the response generation module, where a large language model synthesizes an answer. To ensure factual consistency and mitigate the risk of hallucination—a common issue in generative language models—an additional verification model evaluates whether the generated response is grounded in the retrieved content. A summarization model also condenses interactions to facilitate follow-up queries and maintain continuity in multi-turn conversations.

A dedicated classification mechanism complements the assistant by assigning root cause categories to new incidents. Two approaches have been implemented. The first relies on similarity-based majority voting, where the most frequent class among the top-k similar historical incidents is used as a proxy for classification. The second approach employs a language model-based judge that evaluates candidate classes based on both the incident description and related historical cases. This method offers higher interpretability and flexibility, particularly in complex scenarios with overlapping incident types.

A benchmarking laboratory has been established to assess the quality and robustness of the assistant's performance. This environment includes a standardized dataset of user queries, annotated source documents, and evaluation criteria. The assistant is tested across multiple generations for each prompt to account for variability in model outputs, and performance metrics are monitored via an interactive dashboard interface. This evaluation strategy provides a solid foundation for iterative improvement and comparison between system variants.

Beyond incident investigation, the framework incorporates tools for predictive maintenance by modeling long-term parameter trends. Time series forecasting models have been applied to telemetry data focusing on degradation-prone components, particularly those associated with the Frequency and Timing (F&T) subsystems. Given the limited historical data, forecasting approaches were designed to be lightweight and robust, utilizing global time series models such as Prophet [13]. The Optuna [14] framework enhances the model calibration process, which enables automated hyperparameter tuning. The outputs include trend forecasts and confidence intervals, allowing engineering teams to anticipate parameter deviations and schedule pre-emptive maintenance accordingly.

An integral part of the framework includes anomaly detection tools for investigative and real-time applications. These tools process multivariate telemetry time series using a variety of unsupervised methods. Deep learning-based architectures, such as autoencoders [15] with linear, convolutional, or LSTM-based encoders and decoders, are employed to capture complex temporal dependencies. The reconstruction error, computed between input and output sequences, serves as an anomaly score. Additional methods based on statistical distribution comparison and manifold learning offer alternative detection strategies that do not require GPU acceleration, enhancing the framework's versatility.

The distribution-based model leverages on earth distance metric [16] to compare reference and target time series, while the manifold-based method [17] identifies structural changes in the dimensionality of data representations. Both approaches are beneficial in scenarios where deep learning solutions may be computationally prohibitive. Furthermore, machine learning techniques such as Empirical Mode Decomposition [18] combined with unsupervised clustering algorithms (e.g., Isolation Forest, Local Outlier Factor, see [19] for a comprehensive review) have been incorporated to improve interpretability and reduce false positives. These methods are also compatible with explainability tools like SHAP [10], which help identify the variables contributing most to an anomaly.

Another critical part of the framework consists of pattern-matching techniques. The MASS (Mueen's Algorithm for Similarity Search [21]) algorithm provides a fast distance-based comparison method that can accurately detect candidate matches. Additional statistical tests, including correlation and Kolmogorov-Smirnov analyses, ensure the reliability of the detected patterns. Moreover, a statistical-based approach has been implemented using a theoretical cumulative Gaussian mixture distribution to suggest to the user a set of informative distances, reducing the human intervention during the threshold selection process.

Published by the Canadian Space Agency (CSA) on behalf of SpaceOps, with permission and released to the CSA to publish in all forms."

Dynamic Time Warping (DTW [22]) is used for non-fixed pattern durations to align sequences temporally and account for flexible signal morphology. Pattern matching is achieved through string-based descriptions and Levenshtein distance measures on categorical data.

Another valuable capability of the system lies in its context suggestion feature, which aims to identify telemetry parameters that are statistically correlated with detected anomalies or matched signatures. The system identifies candidate features that may explain or anticipate the anomaly by analyzing the distribution of parameters preceding a signature match and comparing them with unrelated time intervals. A binary classifier is trained on this distributional data, and feature importance is ranked using SHAP values, guiding engineers toward the most relevant telemetry channels for further investigation.

Finally, trend detection tools have been developed to capture slow changes in telemetry parameters that may not be apparent through standard anomaly detection techniques. The methodology involves computing a series of statistical time series from telemetry data and fitting them with models capable of capturing linear and seasonal patterns. The Kwiatkowski-Phillips-Schmidt-Shin (KPSS [23]) test determines the statistical significance of deviations from existing trends, prompting model updates or the segmentation of time periods. This approach allows for a nuanced understanding of parameter evolution over time and provides a complementary layer of monitoring for system health.

These AI tools represent a comprehensive framework for enhancing operational awareness, accelerating incident resolution, and supporting proactive maintenance scheduling in complex technical environments.

5. Results

The structured approach adopted in the Scoping and Ideation phase has laid a strong foundation for AI-driven innovation while maintaining a focus on trustworthiness and usability. By integrating explainable AI methodologies, incorporating continuous user feedback, and proactively managing technical debt, we have positioned the project for a successful transition to operationalization.

To leverage synergies and commonalities across use cases, it was decided to evolve the different MVPs into two core tools: AIIA and SMAI. This approach will enhance operational efficiency, allowing teams to take advantage of overlapping functionalities and create a more cohesive user experience. Sections below present the two merged architectures.

5.1 AIIA - AI Powered Incident Assistant solution

The AI Powered Incident Assistant solution (AIIA) is the operational implementation of the A1 and A2 use cases. As no similar tool existed neither in ESOC or GSOC, the tool is a standalone client, which can be easily integrated with other tool suites as the technology stack is similar (if not the same) to ensure proper maintenance. Nevertheless, the solution interfaces with other systems of the ground station engineers (e.g. OCAI, ARTS, Arrakis) for information traceability and full adoption of the workflow.

Upon login into the application, users will see the list of incidents and the status of their investigations as displayed on Figure 4. Through the overview, they can filter and sort which incidents are more relevant to them and either subscribe or unsubscribe to receive notifications about their updates.



Figure 4: AIIA Incidents & Investigations Overview Page. Users can subscribe for notifications on Investigations, filter by their interest/work area and access the investigations statuses

When they want to check or work on any of them, they can click on the specific row and the system will take them to the detailed investigation as depicted in Figure 5.

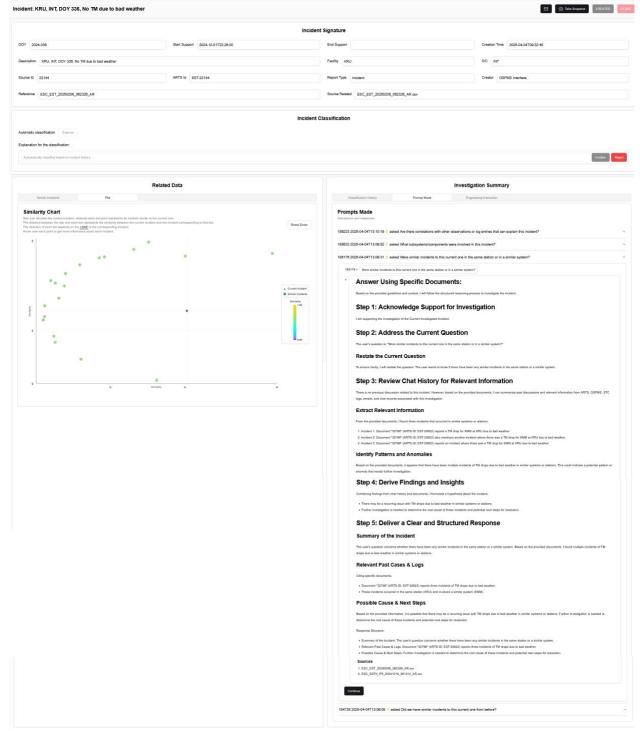


Figure 5: Detail on the Incident Investigation Page. Users can assess the incident information that the engineers record, the automatic classification of the root subsystem, similar past incidents to the one under investigation and all the ongoing interactions including the assistant prompts, and the exchanges between engineers including mails, chat, etc. which become a knowledge repository for future investigations.

SpaceOps-2025, ID #49

Since MVP phase, users were quite adamant the list of similar incidents and a heat map of them were high priority so they could verify their investigations and conclusions instead of trying to investigate hundreds/thousands of incidents. As a consequence, the similarity view is always on display in the solution, and users can switch between tabular view (shown in Figure 6) and a heat plot (see below) to enhance the findings by visualizing the clustering of the similar incidents (e.g. incidents can be similar because of the problem, the cause, the measures taken, etc.) and all previous interactions of the investigation and its status at any time on the top right corner.

Also, security has been addressed during this second phase by adding Role Base Authentication to the solution with actions being limited to the roles and anonymization of the users by the application of roles within each ground station.

However, the core of the solution is Engi, the LLM assistant, which must resemble a knowledgeable colleague who is aware of all past investigations and the human engineer interactions, to provide reasoned responses to the questions of the rest of the engineers.

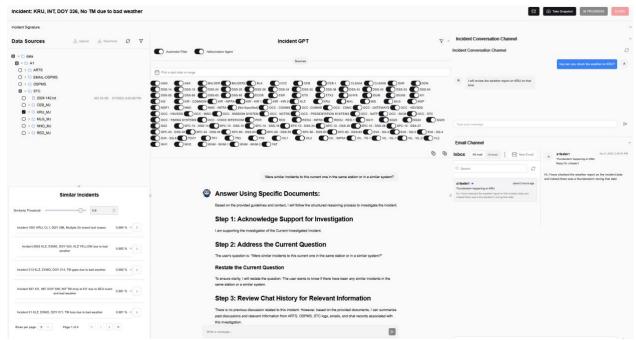


Figure 6: Engi Assistant page. Users can interact with Engi asking questions about the incident and using a range of filters (date, ground station, hallucination enabled, files filtering) and also can view all information available to Engi on the data source panels and the engineering interactions (i.e. chats and emails) related to the investigations that have been provided from the UI itself or from Microsoft365 tool suite used in ESA or by email address itself through IMAP in GSOC.

Although at MVP a first version of Engi was available, during operationalization phase a more robust, performant and complex Agentic System was developed by performing the following improvements:

- Parallelization of the collection grading steps instead of performing a sequential process
- Add a chain of through approach to continue conversations passing summaries of the conversations
- Provide weights to the collections
- Create an optional auto filter for the prompts where the system detects the ground station, spacecraft or time the question refers to.
- Enhancing the user prompts including incident specific information to focus the question
- Includes automatic integration of mails and chats exchanged by Engineers during the investigation dynamically, so results are streaming instead of always looking into the past.

At the moment of writing, all the user interface has already been completely developed through iterations with the stakeholders. However, we are still improving Engi to reduce the hallucination typical of LLMs, offer in parallel narrow and wider perspectives of the same question while test benching on the answers is ongoing by the users not only in ESOC but also by GSOC with the tailoring performed for their specific data collections.

5.2 SMAI - Satellite health Monitoring AI-powered suite

The Satellite health Monitoring AI-powered suite (SMAI) is the operational implementation of the F1 and F3 use cases. It has been integrated into NgWebMUST [24] at ESOC and ViDA [25,26] at GSOC for a wider adoption as both NgWebMUST and ViDA are already extensively used by operators. This approach allows for a smooth integration in the operations workflow. Figure 7 shows how to build a signature by dragging and dropping on-board events, telecommands and telemetry parameters. Figure 8 shows an example of a signature definition in which two events should have happened and a telecommand should not happen. The ruling indicator sets the time and the other indicators time(s) are relative to the ruling one.

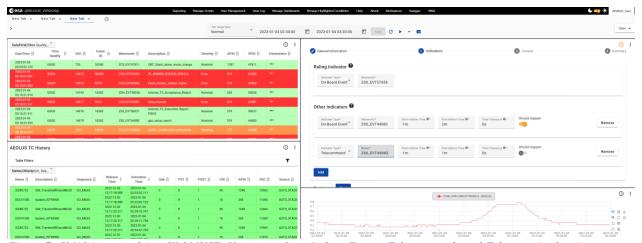


Figure 7: SMAI integrated into WebMUST. Users can drag & drop Events, Telecommands and Telemetry in the signature creation

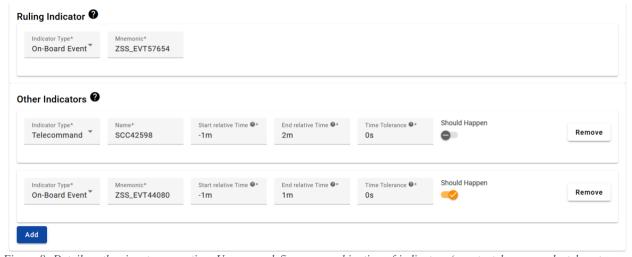


Figure 8: Detail on the signature creation. Users can define any combination of indicators (events, telecommands, telemetry ranges, telemetry pattern matching) and if they should or should not happen.

As part of the operationalization, new features were added to SMAI compared to the MVP phase. These features include monitoring, notifications, reporting, pattern matching, context suggestion, multi-mission support, multi-user support, compatible with multiple data providers.

- Monitoring: if a signature is configured to be monitored, SMAI will attempt to find matches for it when
 new data is downlinked. This process also considers data consolidation to make sure that data is complete
 to provide accurate results.
- *Notifications*: users can decide for which signature matches they want to receive notifications. The notifications are received by email and are also available in a dedicated tab within SMAI.

- Reporting: dashboards allow users to have different representations of signature matches (i.e., histogram, trends, heatmap, statistics, etc.) as different widgets. These representations help users in creating reports on recurrent anomalies and expected nominal behavior.
- Pattern Matching: the signature definition at MVP already supported a TM indicator that allowed operators to specify that a certain TM parameter should be in a user defined range for a minimum period of time to be considered a match. However, this mechanism did not allow to provide more fine-grained details such as the signal should increase, or the relation between several parameters should follow this example. With AI pattern matching, users can provide examples of the behavior they want to consider a match. The pattern matching analysis does not only consider the behavior of individual telemetry parameters but also the relationship between them.
- Context Suggestion: signature matches effectively label a dataset, telling when situations of interest to operators happened (and when they did not happen). The AI context suggestion functionality exploits signature matches to provide operators with suggestions on which other telemetry parameters, events and telecommands could be related to the situation captured by the corresponding signature. This can help operators finding causes for an anomaly and characterizing a known behavior.
- Multi-mission & multi-user support: as part of the operationalization effort, SMAI has been made multi-mission and can support several users simultaneously. Each user can create private signatures for experimentation and share them with their colleagues once they are mature. The notifications are also personalized.
- Multiple data-providers compatibility: to integrate SMAI into ESOC's NgWebMUST and GSOC's ViDA, SMAI must be designed to allow for easy adaptation to data sources. This allows to connect SMAI to other systems by only having to provide a SMAI data access implementation.

At the moment of writing, we are still integrating the Long-Term Monitoring (LTM) capabilities into SMAI. Figure 9 provides an example of a long-term analysis in which linear trends are automatically detected. The times where the long-term trends break is automatically detected. We plan to use the signature mechanism previously described to automatically exclude time periods (e.g., maneuvers, eclipses, etc.) from the trend computation and to enable the possibility of computing trends only when certain conditions are met (i.e., working modes).

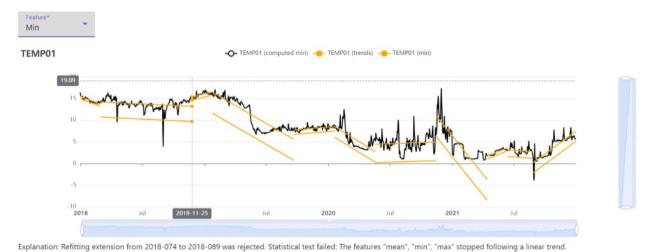


Figure 9: Long Term Monitoring (LTM) at MVP status

The Long-Term Monitoring capability will benefit from all the generic work done in SMAI such as notifications, multi-mission and multi-user support and the multiple data-provider compatibility.

6. Conclusions

In this paper we introduced two tools, AIIA and SMAI, that can play a pivotal role in the on-going transformation of space operations. Together, they lay the foundation for a future where space systems are not only automated but also truly intelligent and collaborative. These innovations pave the way for more sustainable, agile, and autonomous missions.

The integration of Artificial Intelligence into space operations represents a key transformation in how missions will be operated by enhancing system autonomy, optimizing satellite control, and enabling advanced data analytics. As demonstrated across multiple domains, from anomaly detection to space traffic management, AI can empower organizations to act faster, smarter, and more efficiently.

This activity highlights the importance of continued collaboration between space agencies, private industry, and AI researchers. Such a cooperation is essential both to identify and validate viable AI solutions, and to operate satellites in increasingly ambitious and sophisticated ways.

Acknowledgements

The authors would like to thank the teams from Solenix, Starion, Airbus Defence & Space, FBK and GSOC for their work as well as spacecraft and ground stations engineers and operators and infrastructure engineers from ESA's ESOC and DLR's GSOC who provided valuable feedback and support throughout the project. This work was carried out within the framework of the General Support Technology Programme (GSTP) and funded by the European Space Agency (ESA).

References

- [1] G. De Canio, J. Eggleston, J. Fauste, A. M. Palowski, M. Spada, Development of an actionable AI roadmap for automating mission operations, SpaceOps 2023, 17th International Conference on Space Operations, Dubai, United Arab Emirates, 6 10 March 2023.
- [2] G. De Canio, E. Ntagiou, F. Antonello, J. Eggleston, Artificial Intelligence for mission operations automation roadmap: the European Space Operations Centre updates and vision, SpaceOps 2025, 18th International Conference on Space Operations, Montreal, Canada, 26 30 May 2025.
- [3] G. De Canio, K. Annus, Using machine learning for thermal gradients modelling of XMM-Newton propellant tanks, SpaceOps-2021, 16th International Conference on Space Operations, Virtual Edition, 2021, 3 5 May
- [4] G. De Canio, J. Eggleston, A. M. Palowski, M. Schmidt, P. Collins, B. Sousa, D. Mesples, R. Santos, E. Ntagiou, OCAI: the Operations CompAnIon to support decision making of flight control teams, SpaceOps 2023, 17th International Conference on Space Operations, Dubai, United Arab Emirates, 6 10 March 2023
- [5] G. De Canio, J. Eggleston, A. M. Palowski, M. Unal, ESTIM: the ESTrack Investigation and Monitoring tool for analysis of ground station passes, SpaceOps 2023, 17th International Conference on Space Operations, Dubai, United Arab Emirates, 6 10 March 2023
- [6] A. Krstova, P. Fleith, J. Martinez-Heras, N. Salor Moral, A. Buccione, P. Pilgerstorfer, S. Foley, D. Mesples, J. Eggleston, G. De Canio, Intelligent root-cause investigation and AI-assisted handling tool for flight control teams, IAC-24, 75th International Astronautical Congress, Milan, Italy, 2024, 14 18 October
- [7] A. Krstova, F. Hegwein, J. Hansen, P. Fleith, J. Martinez-Heras, J. Lerch, G. De Canio, Artificial Intelligence-based short-term satellite health forecasting, IAC-24, 75th International Astronautical Congress, Milan, Italy, 2024, 14 18 October.
- [8] G. De Canio, J. Eggleston, E. Ntagiou, M. Unal, H. Dreihahn, J. Lerch, M. G.F. Kirsch, S. Foley, N. Salor Moral, P. Pilgerstorfer, M. Szybowski, S. Napoletano, J. Martinez-Heras, A. Krstova, S. Krikorian, Artificial intelligencebased automation of mission post-launch operations processes, IAC-24, 75th International Astronautical Congress, Milan, Italy, 2024, 14 – 18 October
- [9] N. Salor Moral, P. Pilgerstorfer, F. Marino, V. Sicking, M. G. F. Kirsch, A. McDonald, E. Ntagiou, G. De Canio, Automation of flight dynamics planning for ESA's XMM-Newton, IAC-24, 75th International Astronautical Congress, Milan, Italy, 2024, 14 18 October
- [10] G. De Canio, A. Riise, S Mejri, V. D. Tran, I. Tanco, R. Semola, V. De Caro, Advancing satellite health monitoring and control with AI, SpaceOps 2025, 18th International Conference on Space Operations, Montreal, Canada, 26 -30 May 2025
- [11] Grattafiori A. et al., The Llama 3 Herd of Models, (2024) https://arxiv.org/abs/2407.21783
- [12] Dilocker, E. et. al., Weaviate, (2022) https://github.com/weaviate/weaviate
- [13] Taylor, S. and Letham, B., Forecasting at Scale, (2017) https://doi.org/10.7287/peerj.preprints.3190v2
- [14] Akiba, T et. al., Optuna: A Next-generation Hyperparameter Optimization Framework, (2019) https://arxiv.org/abs/1907.10902
- [15] Chen, Z. et.al., Autoencoder-based network anomaly detection, Wireless Telecommunications Symposium, WTS (2018), https://ieeexplore.ieee.org/document/8363930
- [16] Szekely, G. J. and Rizzo, M. L., Testing for Equal Distributions in High Dimension, (2004) https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ad5e91905a85d6f671c04a67779fd1377e86d199

- [17] Chigirev, D. and Bialek, W., Optimal Manifold Representation of Data: An Information Theoretic Approach, (2003) https://www.princeton.edu/~wbialek/our_papers/chigirev+bialek_04.pdf
- [18] Zeiler, A. et.al., Empirical Mode Decomposition an introduction, International Joint Conference on Neural Networks (IJCNN), (2010) https://ieeexplore.ieee.org/document/5596829
- [19] Dobos, D. et.al., A comparative study of anomaly detection methods for gross error detection problems, Computers & Chemical Engineering, (2023) Vol.175.
- [20] Lundberg, S., Lee, S., A Unified Approach to Interpreting Model Predictions, (2017) https://arxiv.org/abs/1705.07874
- [21] Muen, A. et.al. The Fastest Similarity Search Algorithm for Time Series Subsequences under Euclidean Distance, (2022)
- [22] Barrier, Gold, O. and Sharir, M., Dynamic Time Warping and Geometric Edit Distance: Breaking the Quadratic, ACM Transactions on Algorithms, (2018) https://doi.org/10.1145%2F3230734
- [23] Kwiatkowski, D. et.al., Testing the null hypothesis of stationarity against the alternative of a unit root, Journal of Econometrics. 54 (1–3): 159–178. (1992) doi:10.1016/0304-4076(92)90104-Y.
- [24] Silva, José, and Alessandro Donati. "WebMUST evolution." *14th International conference on space operations*. 2016.
- [25] Del Moro, Agnese, et al. "A Modern Approach to Visualise Structured and Unstructured Space Missions' Data." (2023).
- [26] Del Moro, Agnese, et al. "Three (now four) years into operations for a satellite health monitoring system and its evolution" (2025).