Recommendations to Advance Space Trusted Autonomy

Christopher A. Jones,[1] Matthew A. Stafford,[2] Kara Latorella,[3] Christopher Bard,[4] John Dorelli,[5] Erica Rodgers,[6]
National Aeronautics and Space Administration, Washington, DC, 20024, USA

Alejandro Pensado,[7] Gregory Benjamin,[8]
Analytical Mechanics Associates, Inc., Hampton, VA, 23681, USA

Lt. Col Steven Lewis,[9]
United States Space Force, Washington, DC, 20301, USA

Anthony Patrick,[10]
George Mason University, Fairfax, VA, 22030, USA

Jason Hay[11]
Bryce Space and Technology, Alexandria, VA, 22314, USA

The interagency Space Science and Technology Partnership Forum was established in 2015 to identify synergistic efforts and technologies across the U.S. government. While the various space agencies of the U.S. government have distinctly different visions for future operational space systems, all share important foundational common needs. These needs, combined with the maturation of autonomous technology and the prospect of leveraging autonomous systems to address those needs, have led each agency to consider how and when to to implement increasing levels of autonomy in their space systems, and how to determine the trustworthiness of an autonomous system. The Partnership facilitated dialogue among the partners, collected and analyzed data on current and desired future levels of capability, and identified gaps to motivate three recommendations that can be addressed within the Partnership community. These recommendations address the need for more robust documenting and socializing of anomalies in space system operations; the need to expand communication and trust within the community of developers, operators, and end users; and the need for a safe development and testing environment for maturing and demonstrating future autonomous space systems. These recommendations will facilitate both near-term programmatic actions and long-term steps for implementing enduring progress towards enabling space trusted autonomy.

[1] Aerospace Engineer, Space Mission Analysis Branch, NASA LaRC, AIAA Member.
[2] Aerospace Engineer, Space Mission Analysis Branch, NASA LaRC AIAA Member.
[3] Aerospace Engineer, Space Mission Analysis Branch, NASA LaRC.
[4] Research Astrophysicist, Heliophysics Science Division, NASA GSFC.
[5] Research Astrophysicist, Heliophysics Science Division, NASA GSFC.
[6] Science and Technology Partnership Lead, Office of the Chief Technologist.
[7] Aerospace Concepts Engineer, AMA-Inc., AIAA Member.
[8] Aerospace Concepts Engineer, AMA-Inc.
[9] Deputy Chief Scientist, Office of the Chief Scientist, HQ US Space Force, AIAA Member.
[10] Professor Computational Data Science (CDS) Department.
[11] Senior Technologist, AIAA Member.

I. Acronyms

AI = Artificial Intelligence
ARFL = Air Force Research Laboratory
DAF = Department of the Air Force
DRM = Design Reference Mission
HVP = High Value Participant
IAT = Interagency Analysis Team
LOAT = Levels Of Automation Taxonomy
NASA = National Aeronautics and Space Administration
NEO = Near-Earth Object
NPAS = NASA Platform for Autonomous Systems
NRO = National Reconnaissance Office
OCT = Office of the Chief Technologist
S&T = Science and Technology
SMD = Science Mission Directorate
STA = Space Trusted Autonomy

II. Introduction

While the various space agencies of the U.S. federal government have distinctly different visions for future operational space systems, all share important foundational commonalities: improve mission performance, reduce mission costs, incorporate technology advances more rapidly, increase reliability and resiliency, reduce risk, and adapt to anomalies and environmental hazards. These common needs, combined with the maturation of autonomous technology and the prospect of leveraging autonomous systems to address these needs, have led each agency to consider how and when to implement increasing levels of autonomy in their space systems. Simultaneously, the related question of how to determine the trustworthiness of an autonomous system is also escalating in significance.

The Space Science and Technology (S&T) Partnership Forum was established in 2015 to identify synergistic efforts and technologies across the U.S. government, with a focus on key pervasive and game-changing technologies among the space agencies to more efficiently and effectively manage S&T resources [1]. The S&T Partnership consists of three principal agencies: the Department of the Air Force (DAF), the National Aeronautics and Space Administration (NASA), and the National Reconnaissance Office (NRO). Multiple divisions within these three agencies contribute to ensure broad perspectives from each agency. For the Space Trusted Autonomy (STA) topic, the Air Force Research Laboratory (AFRL) contributed under the umbrella of DAF. Because the Partnership membership consists of agency and organization chief scientists and technology officers, it naturally provides a multiagency voice to government-led S&T discussions that advise senior leaders at the principal organizations (DAF, NASA, NRO) on synergies, collaborations, and the state-of-the-art for space technologies.

The S&T Partnership Forum coordinates and facilitates partner dialog, collects data, and performs data analysis to inform science and technology investments. For the STA analysis, the Partnership assembled data products into Recommendations that can be executed within the S&T community at the program and project levels within the contributing agencies and offices. The NASA Office of the Chief Technologist's (OCT) Facilitation and Analysis Team performed these functions for this STA topic on behalf of the USSF Chief Scientist and documented the process and results described in this report. The Facilitation and Analysis Team facilitated discussions with a broad team of subject matter experts from across the Partnership agencies and contributors. The Interagency Analysis Team (IAT), which authored and contributed to this paper, developed an analysis and set of Recommendations for use by the partner agencies. These Recommendations and the process that justifies them are described in this paper.

The S&T Partnership has defined "Space Trusted Autonomy" as:

- "Space" – Focused on space systems
- "Trusted" – Behavior in which human operators and stakeholders have confidence
- "Autonomy" – Some level of decision-making authority that resides within the system

These three terms outline a specialized capability that the Partnership agencies require to overcome physical limitations on future missions (e.g., time, environment). These terms are also the basis for additional definitions in Section III that clarify the nuances among similar terms like autonomy, automation, trust, and trustworthiness.

The IAT systematically developed a set of recommendations related to STA to inform short- and long-term plans for the Partnership. Each of these Recommendations includes a horizon target that will enable the partners to make significant progress towards enhancing existing and enabling new missions. Each Recommendation also includes nearer term steps that can be taken to advance towards that horizon target. Each Recommendation responds to the

IAT input on current state-of-the-art capabilities, a series of envisioned Future States, and the differences between existing capabilities and envisioned future to identify gaps. These differences have been aggregated into eleven Aggregated Gaps, which provide justification for the Recommendations. The Recommendations are also supported by information from the broader NASA community, academia, and the private sector, as elicited through collaborations with outside firms EdgeDweller, yet2!, and NASA's Center of Excellence for Collaborative Innovation.

The first Recommendation addresses the need for more **robust documenting and socializing of anomalies** in space system operations and the use of anomaly data to enable the development of more autonomous capabilities to detect, diagnose, and respond. This Recommendation emerged from observations by the IAT that there are barriers to gathering and sharing the needed information to support development of a variety of capabilities, including those related to handling anomalies.

The second Recommendation addresses the need to **expand communication and trust** within the community of developers, operators, and end users that will work with future autonomous systems. This Recommendation is motivated by several of the Aggregated Gaps that were identified by the IAT. Exercises were conducted to demonstrate the broad applications and issues associated with STA through two elicitations outside the Partnership: one canvassing NASA personnel specifically via the NASA@Work platform, and a directed, global search for related work conducted by yet2!. Further details follow in Section III.G.

The third Recommendation addresses the need for a **safe development and testing environment** in which to mature and demonstrate future autonomous systems. Many of the Aggregated Gaps that were identified by the IAT pointed to this overarching need, as well as to specific near-term steps towards creating that environment. The Recommendation also builds on several visions for future operations in space that were formulated in conjunction with EdgeDweller, an innovation facilitation company that helped develop some of the initial ideas the IAT subsequently used. Discussions with interested industry and academic groups during a government-led technical interchange meeting also pointed to this need.

Section III of this paper describes the process used by the IAT to build on initial discussions with EdgeDweller and during the government-led technical interchange meeting to arrive at the Recommendations. Section IV presents key results from each step of the process. Section V presents the full text of the three Recommendations. Section VI summarizes the effort. An Appendix presents the full text of the Preliminary Gaps that led to the 11 Aggregated Gaps. The Appendix also provides traceability of those Gaps to another framework developed by the IAT and derived from the NASA Autonomous Systems Capability Leadership Team.

C. Process/Methodology

A. Defining Terms

In the initial discussions with EdgeDweller and in the government-led technical interchange meeting, the IAT found that a barrier to collaboration in the Partnership was the lack of common language for defining and discussing the work. A common language fosters understanding across all participants, so less time is spent clarifying concepts and more time can be spent productively developing ideas. As a result of this finding, the IAT compiled thoughts on relevant terms to provide an STA *linga franca*. For the purposes of subsequent Partnership interactions, the following definitions were used:

1) **Autonomy** – Ability of a system to achieve goals while operating independently (or with minimal guidance) of external control with appropriate behavior under known, unknown, and variable conditions.
2) **Automation** – The automatically controlled operation of an apparatus, process, or system by mechanical or electronic devices that take the place of human labor.
3) **Autonomous System** – A combination of elements (e.g., humans and machines) that operates independently (or with minimal guidance) from external control. Note that an autonomous system is not limited to an "uncrewed capability."
4) **System** – A combination of elements that function together to produce the capability required to meet a need. The elements include all hardware, software, equipment, facilities, personnel, processes, and procedures needed for this purpose throughout the lifecycle.
5) **Trust** – A property between an Assessing Agent (Trustee) and a system that is trusted by the Assessing Agent. The property has to do with the confidence the Trustee has that the system/element/etc. will do what the Trustee expects it to do, and not do what the Trustee does not expect it to do.
6) **Trusted** – The system/element has met the confidence threshold of the Assessing Agent (Trustee).

7) **Trustworthy/Assured** – The system/element should meet the confidence threshold of the Assessing Agent (Trustee).

B. Defining Use Cases

Building on the initial discussions from the EdgeDweller facilitation and the government-led technical interchange meeting, the IAT identified a set of Use Cases from their domain of expertise that represented an advancement in STA. The IAT characterized these Use Cases in terms of the following: objective, constraints, performance metrics, human roles, readiness levels, and threats to achieving the envisioned capability. The resulting set of Use Cases (Section IV.A) defined the breadth of considerations for subsequent analyses and served as a starting point for defining the Future States.

C. Defining Future States

Building on the Use Cases, the IAT developed a set of Future States, notionally envisioned in 2035, that would inform capabilities for autonomous operation and the degrees of trust necessary in those operations. The IAT stated the performance objective, identified constituent forms of technology, and conducted a survey to describe the human roles and levels of engagement necessary to achieve that performance objective. Respondents characterized the type (information acquisition, information analysis, decision and action selection, and action implementation) and degree of human interaction for three operational scenarios using the Levels of Automation Taxonomy (LOAT) developed by Save & Feurerberg (Table 1) [2]. The framing of the four task types was sufficiently generic to encompass the breadth of future states under consideration by the IAT, and the levels within each task allowed for characterization of the degree to which automation was used to accomplish each task type. The three operational scenarios were (1) operations during the initial deployment of the system in its operating environment, (2) routine operations in its operating environment, and (3) operations during off-nominal events.

Table 1 Levels of Automation Taxonomy (Save & Feuerberg, 2012)

A INFORMATION ACQUISITION	B INFORMATION ANALYSIS	C DECISION AND ACTION SELECTION	D ACTION IMPLEMENTATION
A0 Manual Info Acquisition	B0 Working Memory Based Info Analysis	C0 Human Decision Making	D0 Manual Action and Control
A1 Artifact-Supported Info Acquisition	B1 Artifact-Supported Info Analysis	C1 Artifact-Supported Decision Making	D1 Artifact-Supported Action Implementation
A2 Low-Level Automation Support of Info Acquisition	B2 Low-Level Automation Support of Info Analysis	C2 Automated Decision Support	D2 Step-by-Step Action Support
A3 Medium-Level Automation Support of Info Acquisition	B3 Medium-Level Automation Support of Info Analysis	C3 Rigid Automated Decision Support	D3 Low-Level Support of Action Sequence Execution
A4 High-Level Automation Support of Info Acquisition	B4 High-Level Automation Support of Info Analysis	C4 Low-Level Automatic Decision Making	D4 High-Level Support of Action Sequence Execution
A5 Full Automation Support of Info Acquisition	B5 Full Automation Support of Info Analysis	C5 High-Level Automatic Decision Making	D5 Low-Level Automation of Action Sequence Execution
		C6 Full Automatic Decision Making	D6 Medium-Level Automation of

			Action Sequence Execution
			D7 High-Level Automation of Action Sequence Execution
			D8 Full Automation of Action Sequence Execution

D. Defining Current Activities

Using the LOAT and the four task types as described for the Future States, the IAT characterized human roles and levels of engagement for each Future State as if the performance objective were to be conducted in the next 12 months (rather than in 2035). The IAT also provided data associated with current activities, funding, facilities, and collaborations that could support each Future State. The resulting descriptions of an envisioned Future State and a Current Activity formed a matched pair that served as the basis for subsequent analysis.

E. Identifying STA Gaps

The performance needs required to progress from Current Capabilities to desired Future States, developed through the survey and follow-on interviews, were defined as Gaps. These interviews probed for technological advancements in performance capability, for robustness to the space operational context, and for approaches to support human trust in these future operations. From these discussions, the IAT developed an initial set of 54 Preliminary Gaps. This initial set included redundant Gaps from multiple defined Future States, as well as Gaps that were sufficiently similar that they could be grouped together. Thus, the IAT refined the initial set of Preliminary Gaps to develop a set of 11 Aggregated Gaps. As an additional product to support subsequent discussions, the IAT also developed a mapping of the Preliminary Gaps to NASA's Autonomous Systems Technology Taxonomy; see the Appendix.

F. Assessing STA Gaps

Leveraging evaluation methods used in other research [3, 4], the IAT developed a Value Model for assessing the Aggregated Gaps. The Value Model gave the Partnership a common basis to assess how addressing each Gap achieves a common goal agreed upon by the principals. The Partnership defined an overarching Goal for the Value Model: evaluate which Gaps best increase efficiency in space mission operations. Efficiency was evaluated through the specific lens of enabling more effective deployment of personnel across satellite operations, achieving near-term improvements on mission operations, and providing a pathway for transformational efficiency for missions to cis-lunar space, the Moon, Mars, and beyond. To respond to this Goal, the Partnership identified five Objectives for evaluating the Aggregated Gaps on five-point Likert scales, to characterize the significance of the Gap for further pursuit:

O1.) **Affordability** – What is the programmatic feasibility of addressing this gap?
O2.) **Technical Feasibility** – What technology maturation is needed to address this gap?
O3.) **Breadth** – How many missions will be impacted by addressing this gap?
O4.) **Immediacy** – How soon will addressing this gap result in impacts in operating space mission architectures?
O5.) **Impact** – To what degree will addressing this gap improve performance for current and future space missions?

Table 2 shows the descriptive labels attached to the levels of 1, 3, and 5 on each Objective's scale; evaluators used 2 and 4 as intermediate values.

Table 2 Gap Analysis Rating Scales

Score Rank	Affordability	Technical Feasibility	Breadth	Immediacy	Impact
1	Programmatic barriers:	Multiple low TRL	Only one mission	More than 5 years from	Modestly enhances

	money alone will not help	capabilities to mature		now	current performance
2					
3	Significant investment required	Some technical maturation required	Some missions	3 to 5 years from now	Significantly enhances current performance
4					
5	Little to modest investment need	Little technical maturation required	Most missions	Less than 3 years from now	Enables new missions

G. Broadened Search

In addition to developing Recommendations from the analysis approach within the Partnership, the FAT conducted two activities with a broader audience. The intent of these two activities was to identify High Value Participants (HVPs) for further discussion. Identifying this broader set of HVPs supports a more complete characterization of the necessary goals of future STA, enabling technologies and methods, and resources (e.g., foci of work, facilities, funding, collaborators) required to achieve the IAT's intended goals.

The first activity was a NASA Agency-wide inquiry through the NASA@Work platform, a crowdsource platform that engages the NASA community. This inquiry was targeted to self-identified individuals who address (1) STA Technology, i.e., work on autonomous space operations enabling technology to reduce dependency on human involvement; (2) STA Design/Test, i.e., work on methods to design or certify complex human, automation, robotic, or autonomous system performance; (3) Human Interface Design, i.e., work on human-system interface requirements to support assessment of complex or remote system trustworthiness; or (4) Operations, i.e., having operated, maintained, or fielded a highly autonomous system. Respondents who identified themselves in one of these roles were then provided follow-up questions based on their self-identified categories. STA Technology respondents were asked to describe their technology applications, detail the aspects of the work considered more automated (using an adaptation of the LOAT, Save & Feurenberg, 2012), and indicate if humans were still involved in the process of using this technology. STA Designers/Testers were asked to classify their methods as basically theoretical or empirical, and then to further describe these methods. Those who address Human Interface Design were asked to describe the tools and methods they use to "enable humans to trust a complex and/or remote system performance prior to fielding." Those who declared themselves associated with Operations of a highly autonomous system were prompted to describe the system and their roles. If individuals were involved in the design of the system and operations, they were asked about what they learned as a result of observing fielding/operations. All respondents had the opportunity, as an alternative to the survey, to provide free text. Survey respondents who permitted follow-up discussion were also asked to provide contact information.

The second activity was similar in intent but identified representative HVPs external to NASA, including industry, academia, and other governmental agencies (foreign and domestic). yet2!, accessed through NASA OCT, specializes in supporting the innovation strategies of organizations, by identifying existing, adjacent, or new technology and market opportunities. The IAT requested that this approach be applied for identifying HVPs for potential future broadening of the STA investigation and as potential future constituents of a broader collaborative partnership. Because yet2!'s methods for searching are proprietary, the IAT provided yet2! with Use Cases from the Partnership and the following topics to seed these searches:

- Technology that supports trustworthiness and trustworthiness assessment of fused data, including uncertainty management;
- Technology that supports reasoning and learning;
- Technology that supports spacecraft/device monitoring, event/anomaly detection, and prediction/prognostication;
- Technology that supports highly autonomous robotic control/actuation;
- Technology that supports multi-attribute dynamic re/planning and scheduling;
- Technology that supports mixed-initiative work (human and synthetic agents);
- Technology that supports distributed communication, coordination, and collaboration among human/synthetic agents;
- Design methods and architectures that enable trustworthy highly autonomous systems;

- Design methods and architectures to establish human/automation and robotics roles and functional allocations (initial, modal, and dynamically);
- Testing facilities and methods (including modeling, simulation, and human-in-the-loop) that assess and provide evidence for trustworthiness;
- Human interface technology and methods to support trustworthiness assessment and calibration during operations;
- Barriers to technology adoption.

Yet2!'s search was limited to publicly available sources and focused on active contributions made between the years 2015 and 2020.

D. Results

A. Use Cases

In scoping the subsequent development of Gaps and Recommendations, a set of Use Cases was developed to characterize potential advancements in specific mission applications. Developed by participants in the IAT across a range of exploration, operations, and technology development domains, these Use Cases served as a context in which the Future States were subsequently formulated. They thus serve two purposes: to explore outside of current paradigms for space missions, and to provide a basis from which the Gaps could be defined. Summaries of each of the seven Use Cases follow.

B. Resource Identification & Mapping in Uncertain Environments

This Use Case envisions a collection of spacecraft with a variety of sensors that provide different coverage, precision, and signature information about the surface of a given planet. The objective is, under a specified time constraint, to survey a planet and map locations of a desired resource. An assumption is that these spacecraft have a communication network that enables information sharing and cooperation between spacecraft platforms. The human-spacecraft interaction involves a team of humans (on Earth) that can interact with the spacecraft via a low data rate, high-latency, intermittent communication link.

C. Detection, Tracking, & Identification of Near-Earth Objects

For this Use Case, a set of terrestrial, space-based, and lunar-based sensor systems provide various sensing modes for detection, tracking, and identification of near-Earth objects (NEOs) that could potentially collide with Earth. These sensor systems have a communication network in place that enables information sharing. The objective of this Use Case is, within a given specified time period, to detect, track, and identify all NEOs that are observable. Autonomy capabilities would allow improved detection, as well as a means for determining how "threatening" the NEO is for Earth.

D. Spacecraft Mission Management Through Anomalous Conditions

This Use Case focuses on maximizing spacecraft mission performance. A single spacecraft is given an orbit with a known mission, but the various environments and anomalies that the spacecraft undergo throughout the mission lifecycle is uncertain (eruptions, solar flares, etc.). The spacecraft adapts to changing conditions to achieve mission objectives; this may include periodic disruptions of communication links to human operators.

E. Autonomous Rendezvous with an Unknown/Uncertain Object: Detection, Approach, Rendezvous, Landing, and Operating on a Near-Earth Object

The objective of this Use Case is to substantially advance autonomy to robustly operate in a vastly uncertain environment: a near-Earth object. Specifically, a spacecraft will approach an NEO and establish situational awareness by scanning the object, determining potential landing sites, and assessing hazards. Once an appropriate site is chosen, the spacecraft lands, gathers surface samples (among other scientific objectives), and returns to Earth.

F. Spacecraft Anomaly Detection, Resolution, and Response

This Use Case focuses on a spacecraft's ability to autonomously detect and resolve anomalies in space. A spacecraft launches into a known orbit, equipped with science instruments capable of measuring the electromagnetic fields and plasma in its immediate environment, as well as an onboard processor capable of commanding the behavior of various spacecraft subsystems. The spacecraft monitors its environment and self-diagnostics for potential anomalies; upon detecting one, the spacecraft determines the cause and takes appropriate action (e.g., initiate safe mode and reboot after the anomaly has passed).

G. Trusted Autonomy for Gateway Enabled by an Artificial Intelligence (AI) Software Platform (NASA Platform for Autonomous Systems – NPAS)

Full trust with autonomous systems has yet to be achieved, and the technology and associated processes for achieving this trust are lacking. This Use Case assumes the NASA Platform for Autonomous Systems (NPAS) will

support autonomy requirements and concepts of operations being established for Gateway and future Artemis systems. The AI software for the autonomy associated with the Gateway is Vehicle System Manager. This AI software applies models and reasoning at high levels of abstraction to allow for onboard "thinking" autonomy that enables affordable, sustainable, and evolutionary autonomy operation—reducing humans in the loop. NPAS autonomy applications encompass use of natural language expressions, which are conducive to achieving AI in an understandable and approachable way and, consequently, enabling trust.

H. On-Orbit Space Manufacturing

Currently, manufacturing and assembling large structures and platforms is done on Earth. Manufacturing large structures on orbit overcomes mass and launch fairing constraints associated with launching large, pre-integrated payloads. On-orbit space manufacturing potentially enables large persistent structures and platforms to be assembled and routinely upgraded to achieve unprecedented technologies and aperture sizes. This manufacturing process requires multiple autonomous systems working in conjunction with on-ground operators and in-space astronauts.

From analyzing the Use Cases, the IAT observed four themes related to STA: Coordination, Management, Learning, and Trustworthiness.

1) Future autonomous systems will need to **coordinate** with multiple actors to produce the desired mission outcome. These actors may include external systems, internal subsystems, and/or human operators.
2) Future autonomous systems will need to **manage** and instruct (sub)systems to respond to both expected and unexpected situations, both internally and within their environment. This will involve diagnosing issues and devising appropriate responses.
3) Future autonomous systems will need to be able to **learn** from and within ever-changing situations to reduce human involvement in directing and planning. This may require the development of a self-management algorithm and/or other supervised, unsupervised, or reinforced learning methods.
4) Finally, to be truly autonomous, these systems need to prove their **trustworthiness**, so human operators may allow them to run with minimal adjustments. This will require advanced validation and verification methods of autonomous decision making. This may also require developing more intuitive methods between humans and systems.

I. Future States

The IAT built on the initial themes and ideas developed in formulating the Use Cases to define five Future States: envisioned capabilities in 2035 that motivate the subsequent definition of Gaps. By developing these future targets, informed by the Use Cases, the IAT could compare the state-of-the-art and thus understand the technical and programmatic Gaps that existed to achieving these desired futures. This section describes findings from the Future State survey, which asked IAT to describe Future States in terms of the following: performance objectives, enabling technologies, and envisioned levels of autonomy/human engagement (when first fielded, in nominal operational conditions, and in off-nominal operational conditions).

The five Future States are summarized below, followed by several findings from the IAT's analysis of them.

A) **Infrastructure for Autonomous Operations** focuses on an architecture and knowledge-server infrastructure (like the World Wide Web) that could be used by autonomous systems to increment their base knowledge and/or to temporarily augment their knowledge about something to perform a task. This concept resembles, for example, how software in information technology devices is currently updated. To accomplish this, updates are pushed through the internet, and the devices acquire new capabilities and improve performance. This Future State also focuses on ontologies that include taxonomies, frameworks, and languages to allow systems to be able to "speak" to other systems and people about what they (the system) are doing/tasked to do and communicate other functions to systems for knowledge management. This Future State addresses the previously identified themes of coordination and learning.
B) **Efficient Crew Size, Response Times, and Human-Machine Trust for Space System Monitoring and Control** concentrates on applying trusted autonomy to routine mission control operator tasks to free up human operators to do other tasks. Routine tasks may include spacecraft state of health monitoring, spacecraft maintenance, resource scheduling, user equipment troubleshooting and help lines, control system performance monitoring, and alarms/warnings/events processing. Specifically mentioned in the survey is a method, or approach, that could be applied to develop, implement, and gain the appropriate levels of trust for spacecraft. The survey response describes this as a "crawl, walk, run" approach, which incrementally implements and slowly increases (based on success from previous trials/tests) the level of autonomy for a

system. Although not explicit, initial steps can include addressing the autonomous capabilities and architecture during the design and development stage via thorough testing and development of the system and the system's autonomous capabilities. Ground tests in laboratories and/or ground testbeds would eventually lead to testing and modifications in real-world environments (with no adversarial presence). By satisfying predefined goals from developers, operators, and stakeholders in each increment, trust with the associated system could be built to increase assurance and confidence in implementing incremental levels of autonomy for the spacecraft for a given mission. This Future State addresses the previously identified themes of management (albeit of routine operations) and trustworthiness.

C) **Spacecraft Anomaly Prediction, Detection, and Isolation** focuses on a future spacecraft's ability to execute its mission independently from external control in a variable and potentially hazardous environment. The spacecraft should be able to independently monitor its local environment, predict probabilities of future hazards impacting the system, and isolate causes of anomalies. This Future State addresses the previously identified themes of management and learning.

D) **Human Machine Teaming** highlights the metacognitive system of interacting, co-dependent humans and machines achieving their goals in challenging, dynamic, and adversarial environments. Steps towards accomplishing the described Future State include when 1) human and machine elements interoperate across roles, units, and hierarchical levels, 2) machine elements seek, share, and transfer knowledge, 3) machine elements assess trust in other machine elements based on a belief about their (the systems') attitudes, motives, and intentions, and 4) human and machine elements seamlessly delegate fair and functional authority and responsibility for decisions and actions. This Future State addresses the previously identified themes of learning and trustworthiness.

E) **Venus Design Reference Mission** discusses how trusted autonomy can be used for NASA's Science Mission Directorate (SMD) medium-term Venus Design Reference Mission (DRM). The key goals for this mission are to survive, detect, communicate, coordinate, and respond using a networked system of autonomous assets (landers and orbiters). Autonomous capabilities will not only reduce mission risks but will also enable and optimize science missions by independently performing operations, with limited to no human communication and interaction, in the harsh environmental conditions of Venus. The team believes that injecting autonomous elements into this mission concept will enable necessary science. Furthermore, the level of autonomy would vary from prescripted automation sequences of actions to increasingly autonomous systems with the ability to ascertain situational awareness, make decisions, and respond robustly to unexpected events. A communication and navigation infrastructure, sensors and controllers, navigation and hazard avoidance systems, and software algorithms are just some of the capabilities that would need to be developed (or enhanced) and implemented to successfully perform this DRM autonomously and with trust. This Future State addresses the themes of coordination and management.

Future State surveys note the importance of having trust in a system for cases where humans are unable to perform tasks or able to provide a delayed response only. In addition to potential latency, Future State E mentions the importance of trust in autonomous systems in extreme and harsh environments. The combination of designing and developing systems that can be trusted in low latency conditions and in harsh environments can ultimately lead to a new level of capabilities and human/system trust, as well as reduce overall mission risk.

Like Future State E, Future State A notes that truly autonomous platforms or systems need to have the capability to share information among each other and/or other systems. (Future State A refers to this information sharing as "thinking autonomy.") This "thinking autonomy" is onboard software that makes decisions for the system, enabling rapid implementation and execution with little to no human input. From most of the Future State survey responses, this "thinking autonomy" is required as missions and systems advance farther away from Earth and perform more complex, complicated, and dangerous missions successfully, as well as enable the trust to do so.

Future State C focused on spacecraft anomaly detection. Like Future State A's reference to "thinking autonomy," Future State C noted that onboard software as a core technology is critical to enabling this STA Future State. This onboard software must have the ability to predict and recognize environmental hazards, recognize when a system is in an anomalous state, determine the cause of the anomalous state, and correctly respond independently to the autonomous state. It is also important to note that the team additionally highlights the need for radiation-tolerant onboard high-performance computing hardware to run this software.

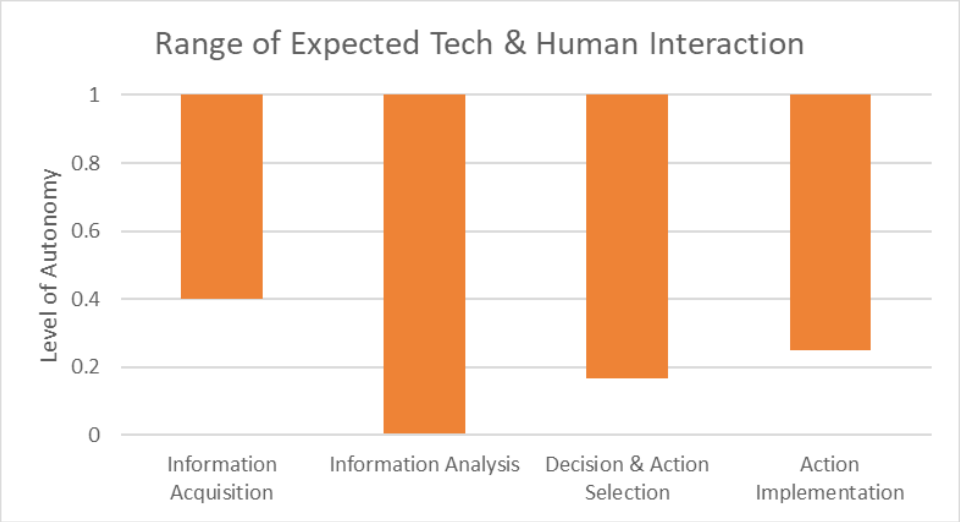


Fig. 1 Normalized Future State responses with respect to the expected interactions between humans and technology. Larger bars represent a greater spread of responses

Based on the results, multiple respondents targeted achieving the highest level of autonomy within the LOAT for all four tasks (see Figure 1), across the operational scenarios described in Section III.C. Information Analysis, the process of involving functions such as memory and inference, had the widest span of scores, based on combinations of scenarios and Future States where that analysis was handled entirely by human operators. All respondents expected a higher level of automation for Off-Nominal operations than First Fielding, potentially indicating a greater confidence in automated capabilities to support handling anomalies than the confidence in those capabilities on initial use. In all three scenarios, the technical capabilities were assumed to be at the same level of maturity; these responses reflect an assumption of lower trust in the first fielding of a system, increased trust during routine operations, and an intermediate level of trust when responding to off-nominal conditions. Respondents identified testing and validation as the most important metric to be trusted early in the lifecycle rather than later in the design lifecycle.

J. Current Activities

For each of the five Future States, the IAT surveyed the current state-of-the-art. The IAT also included how they would evaluate achieving the Future State's performance objectives if a mission or program were started within the next 12 months. In addition to using the results to inform the identification of Gaps (see Section IV.D), respondents identified ongoing projects that would support developments associated with STA, in which nine of the projects were identified with a timeframe of 0 to 2 years, and two projects were identified with a timeframe of 3 to 5 years. No projects were identified for 6 years or more. In mapping these projects to the four LOAT types of tasks, eight were noted as belonging to Decision and Action Implementation, two for Information Analysis, and one for Information Acquisition. Several of the projects were also classified as either a technology development or an internal research and development program, while other projects were not classified by type of investment.

K. Gaps

The IAT used the Future State surveys, Current Activities surveys, and supplementary discussions to develop a set of 54 Preliminary Gaps. As these Gaps emerged from individual discussions, some were redundant with others, and thus the FAT worked to synthesize 11 Aggregated Gaps that encompassed the 54 Preliminary Gaps. The Appendix gives the text of the 54 Preliminary Gaps in Table 8 and the mapping of those 54 to the 11 Aggregated Gaps in Table 9.

In aggregating the Preliminary Gaps, the IAT defined the resulting Aggregated Gap by both what the Gap is and why addressing the Gap is important for achieving the desired Future States described in Section IV.B. The 11 Aggregated Gaps follow:

- **Gap A**: Greater computational power is needed for more autonomous spacecraft capability, because the quantity of data collected by sensors to enable autonomy is too great for existing computational capability.

- **Gap B**: More efficient and comprehensive communication methods and metrics are needed to allow for trust by parties engaging with an autonomous system, because currently there are insufficient data standards and communication protocols robust enough to enable autonomous operations while also providing sufficient cybersecurity to maintain trust.
- **Gap C**: Better standards for defining roles, responsibilities, and requirements between humans and autonomous systems are needed during the design phase, because currently these standards are insufficient or do not exist to enable autonomous spacecraft.
- **Gap D**: Spacecraft need to be able to more quickly detect, diagnose, and respond to anomalies in either mission data (e.g., science data) or housekeeping and subsystem management data, because current spacecraft cannot perform graceful or partial degradations in off-nominal conditions and require significant human oversight to diagnose anomalies.
- **Gap E**: Better automated methods and associated hardware for prioritizing and collecting data are needed, because the increased number and complexity of sources of data will exceed what current operator capabilities can handle.
- **Gap F**: Autonomous decision-making cannot be done in a way that is trusted to meet performance objectives, because there is a lack of platforms that enable analysis and decision-making without humans in the loop, and because there are insufficient ontologies and languages to facilitate implementation of autonomous capabilities, and because those platforms need to be able to explain their decision-making (or have it be inspected) to allow for trust to be built.
- **Gap G**: It is difficult to predict the duration and rigor of testing needed to trust autonomous systems, because there is insufficient knowledge of the time needed to test an autonomous system and of the level of fault tolerance required for systems to operate autonomously for long-duration missions, and because of the lack of capability to test an autonomous system in relevant environments, and because of the need for systems to operate autonomously for long periods of time with little to no communication from ground control.
- **Gap H**: It is difficult to gain experience with and trust in operating a more autonomous system in a contested environment, because there is a lack of testing capability and a lack of procedures for deploying, shaking-down, and upgrading assets in adversarial environments.
- **Gap I**: It is difficult to build trust in autonomous systems, because there is a lack of verification and valid tools and established metrics that can evaluate stochastic decision-making processes.
- **Gap J**: It is difficult to develop common methods that allow spacecraft to autonomously detect and diagnose anomalies, because spacecraft anomaly data is not collected and published publicly.
- **Gap K**: Adoption of new autonomous capabilities by operators takes a long time or does not occur, because there is a lack of coordination between academic and operator communities for trusted autonomy, and because there is a lack of established processes and metrics to use demonstrations to build trust in a system.

L. Value Model

The IAT developed scores for each of the 11 Aggregated Gaps with respect to the 5 Objectives (Affordability, Technical Feasibility, Breadth, Immediacy, and Impact) in the Value Model (see Section III.F.). To do this, the IAT sent a survey that asked the evaluator to rank each gap by the Objective from one to five, with the lowest option as the default. Fifteen subject matter experts from all agencies submitted a response. The results of the evaluation are shown in the following figures (Figures 2 to 6 and Tables 3 to 7). The IAT chose not to aggregate scores, as the purpose of the assessment was not to arrive at a quantitative judgment of each Gap, but rather to foster discussion from different perspectives on each of the Objectives to inform the subsequent development of Recommendations.

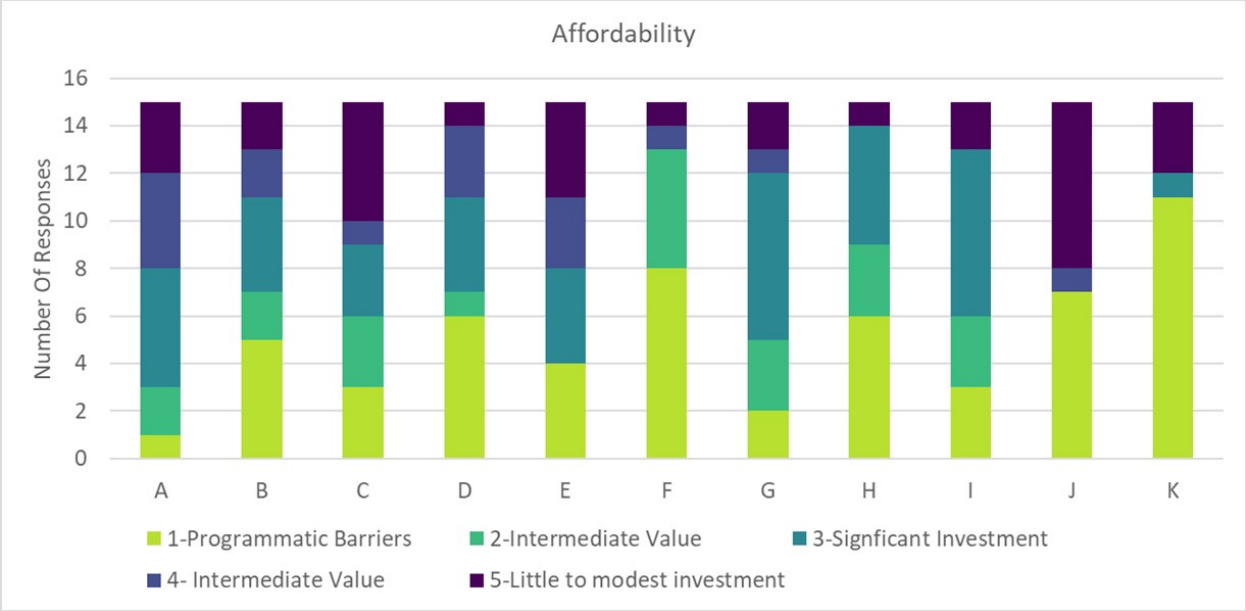


Fig. 2 Aggregated Gaps assessed with respect to Affordability: What is the programmatic feasibility of addressing this Gap?

Table 3 Percent responding to each scoring category for Affordability for each Aggregated Gap

Gaps	Brief Gap Descriptions	1-Programmatic Barriers	2-Intermediate Value	3-Signficant Investment	4-Intermediate Value	5-Little to modest investment
A.	Greater computational power is needed	7%	13%	33%	27%	20%
B.	Efficient & comprehensive communication methods & metrics	33%	13%	27%	13%	13%
C.	Defining roles, responsibilities, & requirements btwn humans and systems	20%	20%	20%	7%	33%
D.	Spacecrafts ability to quickly detect & diagnosis anomalies	40%	7%	27%	20%	7%
E.	Automated methods & hardware to prioritize & collect data	27%	0%	27%	20%	27%
F.	Lack in autonomous decision-making to meet performance objectives	53%	33%	0%	7%	7%
G.	Difficulty predicting the duration & rigor of testing for autonomous systems	13%	20%	47%	7%	13%
H.	Difficulty gaining experience & trust while operating autonomous systems	40%	20%	33%	0%	7%
I.	Difficult building trust in autonomous systems	20%	20%	47%	0%	13%
J.	Difficultly in methods for autonomously detecting & diagnosing anomalies	47%	0%	0%	7%	47%
K.	Operators "learning curve" with new autonomous capabilities	73%	0%	7%	0%	20%

For the Aggregated Gaps, the consensus on a majority of the Gaps was that there are programmatic barriers that would hinder addressing that Gap beyond just increasing funding. **Gap K**, which addresses the "learning curve" required for operators to become proficient at working with autonomous systems, was most frequently identified as being primarily a programmatic issue as this effort encompasses a larger cultural and practical challenge of enabling operators to use new autonomous capabilities. Similarly, **Gap F**, which addresses the need for autonomous decision-making to be done in a way that is trusted to meet performance objectives, is seen largely as a programmatic issue because this perception is more of a cultural issue than a technical one. **Gap G**, difficulty predicting the duration and rigor of testing for autonomous systems, was a Gap predominantly seen as addressed or met through a significant increase in funding.

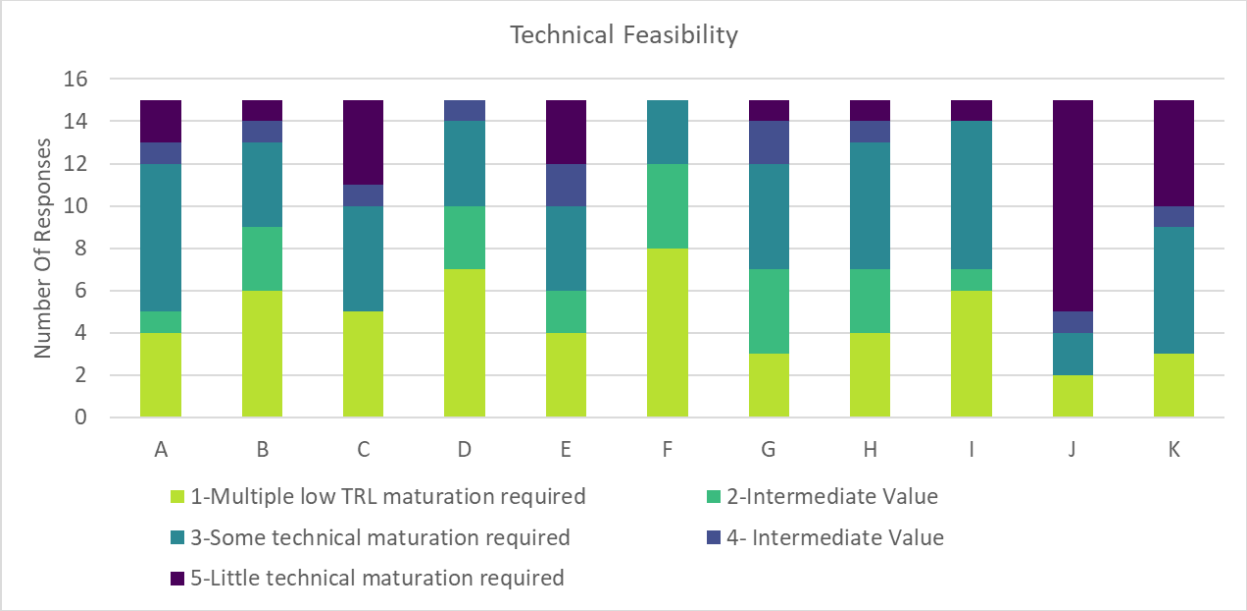


Fig. 3 Aggregated Gaps assessed with respect to Technical Feasibility: What technology maturation is needed to address this Gap?

Table 4 Percent responding to each scoring category for Technical Feasibility for each Aggregated Gap

Gaps	Brief Gap Descriptions	1-Multiple low TRL	2-Intermediate Value	3-Some maturation	4-Intermediate Value	5-Little Maturation
A.	Greater computational power is needed	27%	7%	47%	7%	13%
B.	Efficient & comprehensive communication methods & metrics	40%	20%	27%	7%	7%
C.	Defining roles, responsibilities, & requirements btwn humans and systems	33%	0%	33%	7%	27%
D.	Spacecrafts ability to quickly detect & diagnosis anomalies	47%	20%	27%	7%	0%
E.	Automated methods & hardware to prioritize & collect data	27%	13%	27%	13%	20%
F.	Lack in autonomous decision-making to meet performance objectives	53%	27%	20%	0%	0%
G.	Difficulty predicting the duration & rigor of testing for autonomous systems	20%	27%	33%	13%	7%
H.	Difficulty gaining experience & trust while operating autonomous systems	27%	20%	40%	7%	7%
I.	Difficult building trust in autonomous systems	40%	7%	47%	0%	7%
J.	Difficultly in methods for autonomously detecting & diagnosing anomalies	13%	0%	13%	7%	67%
K.	Operators "learning curve" with new autonomous capabilities	20%	0%	40%	7%	33%

Gap J was the only Gap where a majority of evaluators felt there was little technology maturation needed to address the Gap. From the evaluation, **Gaps A, G, H, I, and K** all need some investment to address technology maturation. **Gaps B, D, and F** were scored by a plurality of evaluators as having multiple low technology readiness level (TRL) technologies that need to be advanced to address the Gap. It is important to note that the top three scores (**Gaps J, K, and F**) need a coordinated effort to meet the gaps, in terms of developing common methods, establishing roles, and implementing capabilities for operators to develop a relationship with autonomous systems.

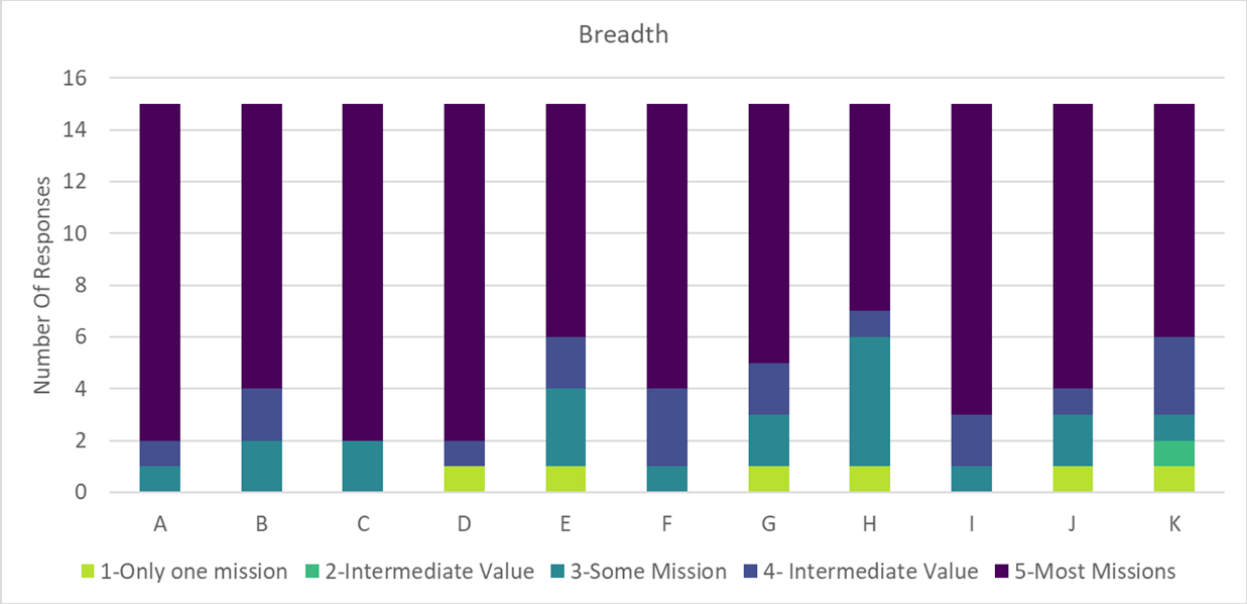


Fig. 4 Aggregated Gaps assessed with respect to Breadth: How many missions will be impacted by addressing this Gap?

Table 5 Percent responding to each scoring category for Breadth for each Aggregated Gap

Gaps	Brief Gap Descriptions	1-Only one mission	2-Intermediate Value	3-Some Missions	4-Intermediate Value	5-Most Missions
A.	Greater computational power is needed	0%	0%	7%	7%	87%
B.	Efficient & comprehensive communication methods & metrics	0%	0%	13%	13%	73%
C.	Defining roles, responsibilities, & requirements btwn humans and systems	0%	0%	13%	0%	87%
D.	Spacecrafts ability to quickly detect & diagnosis anomalies	7%	0%	0%	7%	87%
E.	Automated methods & hardware to prioritize & collect data	7%	0%	20%	13%	60%
F.	Lack in autonomous decision-making to meet performance objectives	0%	0%	7%	20%	73%
G.	Difficulty predicting the duration & rigor of testing for autonomous systems	7%	0%	13%	13%	67%
H.	Difficulty gaining experience & trust while operating autonomous systems	7%	0%	33%	7%	53%
I.	Difficult building trust in autonomous systems	0%	0%	7%	13%	80%
J.	Difficultly in methods for autonomously detecting & diagnosing anomalies	7%	0%	13%	7%	73%
K.	Operators "learning curve" with new autonomous capabilities	7%	7%	7%	20%	60%

Of the five Objectives, Breadth had the strongest consensus amongst all the respondents: most of the Aggregated Gaps apply to most or all envisioned missions. This motivates two findings: first, the Aggregated Gaps developed in this process are sufficiently universal to warrant cross-agency attention; and second, the Recommendations developed from this assessment should be applicable to a wide range of missions, rather than any single application. To that end, the Recommendations in Section V are not specific to a given Use Case or Future State but are instead broad enough to impact missions throughout the cross-agency portfolio.

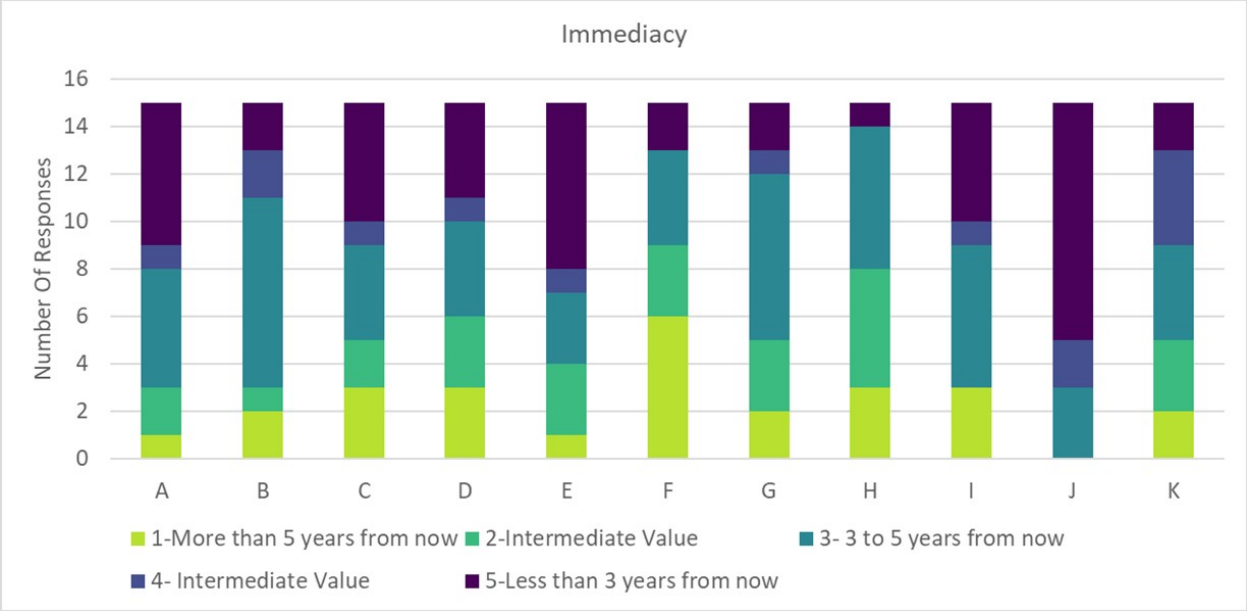


Fig. 5 Aggregated Gaps assessed with respect to Immediacy: How soon will addressing the Gap result in impacts in operating space mission architectures?

Table 6 Percent responding to each scoring category for Immediacy for each Aggregated Gap

Gaps	Brief Gap Descriptions	1-More Than 5 yrs From Now	2-Intermediate Value	3-3-5 yrs From Now	4-Intermediate Value	5-Less Than 3 yrs From Now
A.	Greater computational power is needed	7%	13%	33%	7%	40%
B.	Efficient & comprehensive communication methods & metrics	13%	7%	53%	13%	13%
C.	Defining roles, responsibilities, & requirements btwn humans and systems	20%	13%	27%	7%	33%
D.	Spacecrafts ability to quickly detect & diagnosis anomalies	20%	20%	27%	7%	27%
E.	Automated methods & hardware to prioritize & collect data	7%	20%	20%	7%	47%
F.	Lack in autonomous decision-making to meet performance objectives	40%	20%	27%	0%	13%
G.	Difficulty predicting the duration & rigor of testing for autonomous systems	13%	20%	47%	7%	13%
H.	Difficulty gaining experience & trust while operating autonomous systems	20%	33%	40%	0%	7%
I.	Difficult building trust in autonomous systems	20%	0%	40%	7%	33%
J.	Difficultly in methods for autonomously detecting & diagnosing anomalies	0%	0%	20%	13%	67%
K.	Operators "learning curve" with new autonomous capabilities	13%	20%	27%	27%	13%

The Immediacy Objective addresses the question associated with how soon the impacts will be seen by closing these Gaps. A plurality of evaluators found that for **Gaps A, C, E, and J**, the impacts to the space mission architectures were expected to be seen within 3 years from now if the Gaps were addressed. In the middle of the scoring range, **Gaps B, G, H, and I** had the most scores for having an impact of 3 to 5 years on space mission architectures if addressed now. **Gap F** was the only Gap to score highly with the expectation that it would take more than 5 years to impact space mission architectures. As this Gap addresses human trust with autonomous systems, and the system's ability to autonomously make correct decisions, it will take a significant amount of time for operators and stakeholders to build trust in an autonomous system capability to successfully perform mission objectives.

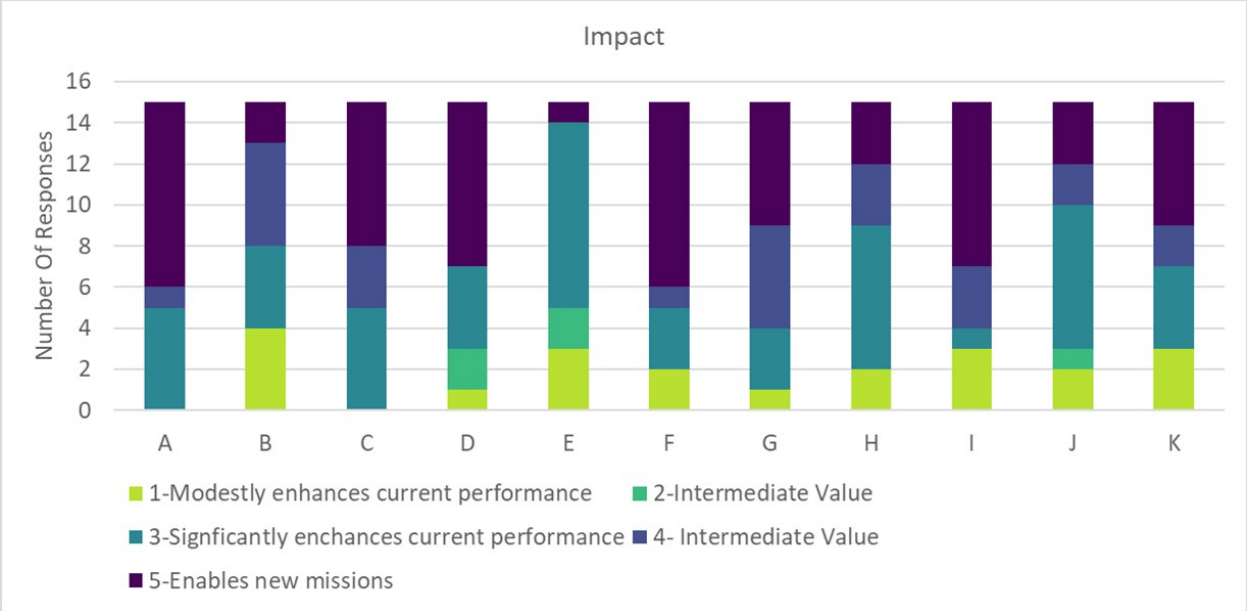


Fig. 6 Aggregated Gaps assessed with respect to Impact: To what degree will addressing this Gap improve performance for current and future space missions?

Table 7 Percent responding to each scoring category for Impact for each Aggregated Gap

Gaps	Brief Gap Descriptions	1-Modestly enhances	2-Intermediate Value	3-Significantly Enhances	4-Intermediate Value	5-Enables New Mission
A.	Greater computational power is needed	0%	0%	33%	7%	60%
B.	Efficient & comprehensive communication methods & metrics	27%	0%	27%	33%	13%
C.	Defining roles, responsibilities, & requirements btwn humans and systems	0%	0%	33%	20%	47%
D.	Spacecrafts ability to quickly detect & diagnosis anomalies	7%	13%	27%	0%	53%
E.	Automated methods & hardware to prioritize & collect data	20%	13%	60%	0%	7%
F.	Lack in autonomous decision-making to meet performance objectives	13%	0%	20%	7%	60%
G.	Difficulty predicting the duration & rigor of testing for autonomous systems	7%	0%	20%	33%	40%
H.	Difficulty gaining experience & trust while operating autonomous systems	13%	0%	47%	20%	20%
I.	Difficult building trust in autonomous systems	20%	0%	7%	20%	53%
J.	Difficultly in methods for autonomously detecting & diagnosing anomalies	13%	7%	47%	13%	20%
K.	Operators "learning curve" with new autonomous capabilities	20%	0%	27%	13%	40%

Lastly, the Impact Objective addresses the degree to which any of the Gaps will improve performance for current and future space missions. From the data above, the table indicates that many respondents found that addressing the STA gaps have impacts that enable new missions and significantly enhance current mission performance. Some evaluators remarked that these represent two distinct Objectives, and they thus used the four score to indicate where they believed addressing a Gap would address both; this was discovered during discussions around the scoring for **Gaps B and G**. A majority of respondents felt that addressing **Gap E** would significantly enhance current mission performance without having a meaningful impact in enabling new missions; from discussions, the FAT determined that evaluators viewed this Gap as representing an improvement over existing capabilities rather than a novel advance, which correspondingly motivated the resulting scoring.

Gaps A, C, and J scored high across all five Objectives. In general, the evaluators noted that addressing these gaps would enhance or enable new missions, would have a relatively quick impact, have a wide breadth, need some or little technical maturation, and require significant to little investment. Gap A is a technical gap that would open opportunities to expand Trusted Autonomy. Gaps C and J are programmatic gaps that could significantly enable how the agencies tackle trusted autonomy in a more unified front.

M. Broadened Search

There were two exercises to broaden the use cases and search for HVPs working in areas that would address STA. The NASA@Work effort sought these from the NASA community, and the yet2! effort sought these more broadly.

The NASA@Work call received 24 responses from 7 Centers, who then self-sorted into the four categories defined by the first survey question: STA Technology, STA Designers/Testers, Human Interface Design, or Operations.

The nine *STA Technology* respondents provided these 10 application areas:

1) Automated communications for spacecraft
2) Autonomous, modular robotics system designed to perform in-space assembly
3) Integrated system health management for sustainable habitats
4) Advancing xEMU display ECI enabling crew to be less reliant on ground support during EVAs
5) Autonomous components for developing large-scale spacecraft structures
6) Sustainable power for spaceflight without humans in the loop
7) Autonomous rendezvous and capture of a non-cooperative satellite
8) Demonstrate space assembly and manufacturing (3D printing)
9) Robotic systems; e.g., enable Landsat7 servicing, and Mars Sample Return Capture Containment and Return System to return Mars surface samples
10) Real-time integration of human physiological sensors, Dynamic Function Allocation (DFA) protocols, and machine learning algorithms to detect and identify suboptimal mental states and facilitate machine/human coordination

This demonstration effort, even with the small sample size, reveals the breadth of STA applications: from autonomous subsystems management (power, communication), to the integrated health management of a complex system, to robotic design, planning, and manual operations. Response 4 underscored that the term "autonomous" is often interpreted with respect to ground operations, so technologies that permit crew to be less reliant on ground are considered STA. Response 10 underscores, similarly, that STA may include increased levels of human-independent "human autonomous" capabilities but does not preclude effective use of humans to maximize performance. As such, technological advancements to human/machine interfaces is also relevant to advancing STA. This point is underscored by the 70% of STA Technology developers who acknowledged that they have a role in their STA application.

Five respondents identified as *STA Designers/Testers*: three characterized their approaches as primarily theoretical (formal methods, architectural frameworks), and two characterized their approaches as primarily empirical (testing, data-based modeling/simulation). These responses again represent a sense of the breadth of approaches available for instilling trustworthy performance and ascertaining performance of STA-enabling technologies. The types of responses provided are as follows:

- Empirical: Simulator with different levels of fidelity, run Monte Carlo simulations, use post-processing scripts to identify requirement violations.
- Empirical: Black box testing of Space Launch Systems Flight Software.
- Theoretical: Research in runtime verification (RV) framework for specifying properties using different logics and then generating monitors from those specifications.
- Theoretical: Deep space systems health management architecture design to dynamically allocate autonomous control systems in a deep space habitat among space vehicle, onboard crew, and ground control.
- Theoretical: Formal methods (theorem proving and model checking) to generate artifacts for safety arguments to justify safety claims.

Human Interface Designer respondents were asked to describe the tools and methods they use to "enable humans to trust a complex and/or remote system performance prior to fielding." Trustworthiness necessitates both the possession of the inherent qualities necessary for a system to perform as intended, and for being able to convey confidence in those qualities to the humans evaluating that trustworthiness. As such, it is crucial to ensure that the Human Interface to STA capabilities permits the effective assessment of their trustworthiness, in design and during operations, and supports the level of involvement of human operators to ensure best holistic performance in all operating scenarios. Comments describing this kind of work at NASA included references to designating equipment to run without human intervention, focusing on formal assurance methods to apply to different phases of the development life cycle of software, creating a theoretical framework for calibrating user trust in automated systems, and increasing transparency to appropriately calibrate trust of AI-based systems.

While some of these responses were not to the point of the question, and rather explained the specific software tools used or degree of human autonomy of their equipment, the latter three comments demonstrate the type of work necessary to advance STA to a point where humans must still ascertain trustworthiness.

Operators, Maintainers, or Fielders of a highly autonomous system were prompted to describe the system and their roles. All three of those who selected this role further indicated that they were designers of system operations (rather than operators or maintainers), and all indicated that they were involved in fielding and post-fielding operations. These survey respondents were asked "As a designer who was involved in system fielding and/or operations, what did you learn about human trust of complex systems from this experience?"

The three received responses underscore three significant issues associated with human trust in complex, highly autonomous systems. These three issues are the necessity of evidence to support trust and the two forms of mistakes when trust is not, or is inappropriately, assessed: over-trusting, and the risks associated with system performance failures, and under-trusting, and the risks associated with failure to use technology.

In the second activity, *yet2!* identified persons whose work contributed to the key areas provided by the FAT. yet2! reviewed 65 potential solutions and presented 38 to NASA. NASA identified 33 solutions that represented these key areas, focusing on aerospace applications. After reviewing identified results, yet2! recommended four additional terms that might be useful in future searches: "Calibrated Trust and Behavior," "XAI" (Explainable AI), "Verifiable AI," and "Autonomous Mitigation." yet2! provided a summary page for each of the identified solutions and a compilation of references obtained.

E. Recommendations

The IAT reviewed the Gap assessment activity, as well as the data and discussions that led to it, to develop initial drafts of three cross-cutting Recommendations that addressed a majority of the Gaps. Through subsequent discussions among the Partnership, these Recommendations were revised to include both a far-reaching vision, as well as one or more near-term steps that could be taken by members of the Partnership to advance STA. Structured this way, each Recommendation carries both a Programmatic component (the specific step that can be taken to have an easily realizable impact) and an Enduring component (the horizon-scale vision that is expected to have high impact across a wide breadth of future missions, including enabling new missions). The text of the three Recommendations, and accompanying commentary, are presented in the following sections.

A. Subsection: Anomalies

From the IAT's assessment of the Aggregated Gaps, **Gap J** stood out as being considered short-term, broad, and feasible to address from both the affordability and technical feasibility Objectives. Similarly, **Gap D** also stood out as having a high impact across many potential missions. Interestingly, **both J and D** scored high on programmatic barriers, indicating the need for an organized effort on anomaly detection and handling. The first Recommendation addresses both Gaps.

> **The S&T Partnership recommends that future missions require the storage of spacecraft anomaly data and metadata to enable development of capabilities to more quickly detect, diagnose, and respond to anomalies. In addition, the S&T Partnership recommends a pilot program to form a consortium of existing missions to document and share satellite operations data among groups. This pilot program will create a framework for storing relevant data, including telemetry, housekeeping, anomaly, human annotation, and other appropriate metadata. This data will be shared among the constituent members. This pilot program could be the first step towards a broader consortium of spacecraft manufacturers and operators across government and industry dedicated to a data strategy that is discoverable and accessible to enhance future missions. The nature of what data is collected and shared will have to be determined by the consortium in the context of barriers and reluctance to sharing data across different manufacturers.**

B. Subsection: Establishing Operator Trust within the Community

Three Gaps pointed to the need to improve the literacy and integration among stakeholders in developing, implementing, and using autonomous capabilities: **Gaps C, I, and K**. The second Recommendation addresses these three Gaps.

The S&T Partnership recommends taking steps towards a more autonomy-literate community of developers, operators, and end-users. Toward that end, the S&T Partnership recommends that existing programs for operator education should be expanded to enhance current relationships among data end users, mission operators, and developers with the goal of addressing trust concerns in newer autonomy systems. This expansion should be based on a review of developer attempts to improve technology adoption and soliciting operators for first-hand experiences. Furthermore, these programs should leverage the experience and lessons learned by data end users, mission operators from different agencies, and developers. Development of autonomy capabilities must include all stakeholders in the design process. For example, knowledge capture from experienced operators will enable more rapid and effective implementation of capabilities.

C. Subsection: Development and Testing Environment

Many of the Aggregated Gaps (**Gaps C, E, F, G, H, and I**) indicate that beyond the specific technologies that must be matured to enable greater degrees of trust in spacecraft autonomy, there is a larger gap in the ability to develop and test capabilities and assess their trustworthiness. The third Recommendation addresses these six Gaps.

The S&T Partnership recommends the creation and development of safe and benign development and testing environments, including one or more in-space testbeds and associated logistics and ground infrastructure, to allow for testing autonomous capabilities and the verification and validation approaches used to assess and calibrate system trustworthiness of those capabilities. Such environments should be created such that the three partners can contribute resources and expertise as their individual agencies allow. As an initial step, we also recommend a series of surveys be conducted to inform the development of those environments:

- **Identify current assets, near end of life, in a space environment that could be leveraged for autonomy testing without compromising their mission.**
- **Identify modeling and simulation capabilities, high-fidelity virtual testbeds, and testing facilities among the agencies that can be used to leverage testing of autonomous systems on the ground to build trust between the system and operators—eventually leading to the testing of these systems in space.**
- **Identify comprehensive methods and metrics to characterize human and autonomous system interactions and the dynamics of trust.**
- **Identify verification and validation tools, metrics, and platforms that can be leveraged to evaluate intelligent reasoning and decision-making processes.**

F. Summary

The IAT has developed a set of Recommendations for the S&T Partnership to use with their member agencies to respond to Gaps in STA identified by the S&T Partnership. The IAT developed desired Future States (derived in part from a series of potential Use Cases) that would use STA in the circa 2035 timeframe. The IAT also developed Current Activities that could be used to better understand the state-of-the-art of relevant capabilities. The IAT examined the differences between the Future States and Current Activities and identified 54 Preliminary Gaps, from which the IAT synthesized 11 Aggregated Gaps. The IAT developed and implemented a Value Model to assess those 11 Aggregated Gaps against 5 Objectives that responded to technical and programmatic needs relevant to the Partnership. From the resulting assessment, the Partnership developed 3 Recommendations, focusing on the documenting and socializing of anomalies, the expansion of communication and trust in communities that work in STA, and the creation of safe development and testing environments for future autonomous systems. These Recommendations will facilitate both near-term, programmatic actions and long-term steps for implementing enduring progress towards enabling STA.

Appendix

Table 8 The 54 Preliminary Gaps derived from the Future States, Current Activities, and Use Cases

Gap ID	Gap Description
1	There is a need to establish cybersecurity for autonomy to trust information coming from autonomous systems.

2	Need procedure(s) to deploy/shakedown/upgrade assets in a contested environment.
3	There is a lack of testing and computational capability required to meet Future States.
4	Although automation technology exists to perform some operator functions, autonomy that is intuitive enough to recognize and react to what an active adversary is doing does not currently exist.
5	Currently, there is very little ability (and quite frankly trust) for spacecraft's to make decisions on their own without a human in the loop, which would allow for better onboard decision making to respond to anomalous situations.
6	The software that enables spacecraft to autonomously respond to computed inferences will need to be tested at higher levels of fault tolerance and in ways that address known unknowns and unknown unknowns.
7	Current spacecraft cannot perform graceful/partial degradation to maximize performance in off-nominal conditions.
8	Currently, isolating the cause of an anomaly requires a group of human experts guided by a set of mathematical rules (e.g., Failure Mode and Effects Analysis, Fault Tree Analysis).
9	Currently, there are only nascent software platforms that enable analysis and decision making onboard (e.g. NPAS), by using physics or other models that experts use to provide insight and solutions.
10	The computational performance of current sensor technology is not developed enough to meet the demands of future autonomy.
11	As the number, complexity, and location of assets increase, processing of data from many sources will surpass operator capacity to handle.
12	Software platforms to implement onboard trusted "thinking" autonomy.
13	Increasing the ability of spacecraft to respond autonomously will necessitate increased computational power and its associated impacts on the spacecraft design.
14	There is a need to develop a natural (intuitive) language for autonomous systems to communicate with humans and vice versa.
15	Ontologies and languages for comprehensive systems models and for implementation of autonomy capabilities.
16	There is a lack of the sophisticated sensor technology required for future autonomous systems (a sensor that can make decisions of data collection without ground support).
17	Greater automation, and likely autonomy, is needed to cover times of no communication with assets.
18	Autonomous systems need to be trusted to operate properly during dormancy for long periods of time without relying on ground control for communication.
19	For space science data acquisition systems, future needs require faster diagnosis of and response to onboard performance anomalies (this may include significant events in the science data—but the housekeeping data and system management are of primary concern); this response must include subsystem management and possible reconfiguration.
20	Spacecraft need to detect and predict anomalies to best preserve themselves and maximize science objectives without unnecessary human involvement.
21	More operations testing is required to determine nominal and off-nominal operating conditions.
22	Need facilities and resources to have operators train and practice operations in a contested environment with the current and planned automated/autonomous system.
23	Successful demonstrations that prove requirements and capabilities are needed to enhance operators' and stakeholders' trust in the system.
24	Testing of autonomous systems in extreme environments is required to increase subsystem/system reliability.
25	Currently, there are no research or testing methods that predict the amount of time needed to test autonomous systems to determine if they have the necessary capabilities for long-duration missions (e.g., health management testing).
26	The DSN performance, capability, and reliability infrastructure is not there for Future State demands.
27	There is a lack of understanding of autonomous systems' reasoning and decision-making process.

28	Machine learning must be viewed as a “grey” or “white” box, rather than a black box, for DoD to implement in strategic/national assets.
29	Paradigms to implement onboard trusted “thinking” autonomy.
30	There need to be established standards defining roles and responsibilities between humans and autonomous systems.
31	There is a need to have a means of coordinated data exchange (communication protocol/data standard) trusted by all parties.
32	No requirements for automated spacecraft control currently exist.
33	No internationally recognized standards for collision avoidance among satellites currently exist.
34	There is a lack of understanding regarding how non-linear/chaotic systems may scale up.
35	New DoD object-tracking capability needed in order to address increasing cis-lunar operations due to the complications of 3-body physics.
36	Metrics do not exist to certify autonomy capabilities (e.g., machine learning) in the challenges of deep space environments.
37	Verification & Validation needs to develop as new autonomy, such as machine learning, is implemented.
38	No agreeable process exists to build stakeholder confidence to move from the current “zero-to-very low” level of automation for satellite operation to the desired “low & mid” levels of automation.
39	Verification & Validation processes are not mature enough to build the “initial” trust from decision makers.
40	Verification and Validation tools for building trust in autonomous capabilities.
41	Spacecraft anomaly data is not collected and published publicly.
42	There are gaps in ongoing work and experience between the academic and operator communities.
43	High volumes of spacecraft data relative to current human staffing and capabilities make it difficult to notice patterns, correctly distinguish off-nominal events from noise, and identify interesting events in a timely manner, all without introducing biases into the interpretation.
44	The utility of crowdsourcing pattern recognition to facilitate data interpretation is limited by the complexity of the problem and the training required to interpret the data.
45	Machine learning algorithms that can operate on small data volumes are needed.
46	There is a lack of automated methods for prioritizing and collecting the most scientifically valuable samples/data.
47	“Where is the gap in someone’s ability to trust autonomy?” is a key question since “trust” or “trustworthy” likely varies by audience.
48	Ability to accommodate in-flight or relevant environment testing (in addition to technology).
49	Formal Verification & Validation methods need to be drafted as testing alone is insufficient. How can a probabilistic decision-making process be Verified & Validated?
50	Look to how deep space missions leveraged autonomy and how those solutions can be repurposed in other domains. Development of “Task Networks.”
51	Accurate assessment of “planner/scheduler” capabilities to determine the level of need should increase trust in the system.
52	Lack of the needed amount and type of sensors to give an autonomous system enough SA to make independent decisions without causing harm. Needs to diagnose and understand its own current state.
53	System development process should allow for rapid testing of (virtual?) solutions to get to unconventional solutions. MBSE should be leveraged heavily.
54	Viable solutions to problems will still exist outside of neural nets and machine learning.

The IAT organized the 54 Preliminary Gaps into subcategories within the four major categories of the Autonomous Systems Taxonomy defined by the NASA Autonomous Systems Capability Leadership Team (AS-CLT) in 2018:

- Situation & Self-Awareness (SSA)
- Reasoning and Acting (R&A)
- Collaboration and Interaction (C&I)

- Engineering and Integrity (E&I).

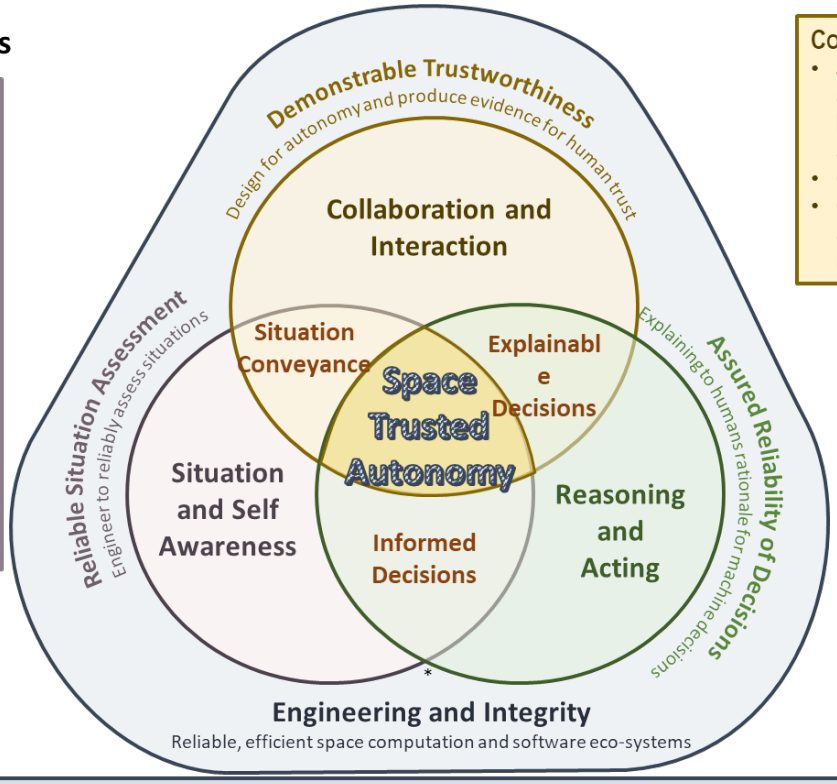
This initial classification enabled the development of a refined classification scheme based on the Autonomous Systems Taxonomy (Fig. 7) that identified cross-cutting areas.

Mapping S&T Original Gaps

SITUATION AND SELF AWARENESS

- **Sensing and perception:** adequate, smart, and adaptable sensing for decision making and diagnosis [10,16,35,52]
- **State estimation:** retention of data provenance and uncertainty conveyance in fused data
- **Knowledge and model building:** modeling environment, system, and objectives [15]
- **Hazard assessment:** among systems and between systems and environments (e.g. satellite collision hazard) [33]
- **Event and trend identification:** unbiased event detection (anomalies, patterns) in large data volumes [43,46]

[#] Gap ID



COLLABORATION AND INTERACTION

- **Joint knowledge and understanding:** mutual situation awareness among humans and machines through natural-language and visual communication, coordinated data exchange, and interrogatable and explainable systems [4,11,14,31]
- **Goal and task negotiation** among humans and machines [30]
- **Operational trust building:** identifying demonstration features and metrics to characterize the dynamics of trustworthiness in machines for operators [5,12,23,29]

REASONING AND ACTING

- **Planning:** efficient and scalable to more complex physical systems, traceable decisions, integrated model of mission prioritization [34]
- **Execution and control:** traceable and understandable decision making; adaptation to physical changes; execution during communication blackouts; human intervention per level of engagement [9,10,17,27,32]
- **Fault diagnosis and prognosis:** tools for onboard diagnosis and predicting faults for self-preservation [8,20]
- **Fault response:** graceful degradation following anomalies; faster diagnosis and response for self-preservation and opportunistic science [7,19]
- **Learning and adapting:** efficient learning from sparse and historical data; responsive to environmental and system functionality challenges and changes in mission objectives [45]

ENGINEERING AND INTEGRITY

- **Design and architecture:** ontologies/languages for modeling; white/grey-box designs; performance assessment to inform design; computation; curation and utilization of historical data; international standards [13,28,41,51]
- **Modeling and simulation:** testing prior to hardware availability and/or when relevant environment is unavailable (including interfacing environment, systems, and humans) [53]
- **Test and evaluation:** procedures, facilities, resources for testing representing contested and challenging environments including human roles; assessing scope and adequacy of testing in relevant environment under nominal/off-nominal and known unknowns and unknown unknowns [2,3,6,21,22,24,25,26,46]
- **Verification and validation** of decision-making and learning systems to provide trustworthiness evidence for stakeholders [36,37,39,40,49]
- **Operational assurance:** a process for stakeholder and operator assessment of trustworthiness for human/machine systems and for dynamically adjusting autonomy levels [1,18,38]

Fig. 7 S&T Preliminary Gaps mapped to Refined Classification Scheme

Table 9 Aggregated Gaps mapped to gaps (X) and subsumed gaps (S)

Prelim Gap ID	A	B	C	D	E	F	G	H	I	J	K
1		X									
2								X			
3	S						S				
4								X			
5						S					
6							X				
7				S							
8				S							
9						X					
10	S										
11					X						
12						S					
13	X										
14		X									
15		S				S					
16					S						
17											
18							X				
19				X							
20				X							
21							S				
22								S			
23											S
24							X				
25							S				
26											
27						X					
28						S					
29						S					
30			X								
31		X	S			S					
32											
33											
34					X						
35											
36									S		
37									S		
38											X
39											S
40									X		

41										X	
42											X
43					S						
44					S						
45					S						
46					X						
47									S		
48							S				
49									S		
50							S				
51											S
52				S							
53							S				
54					S						

Table 10 Aggregated Gaps Mapped to Recommendations

Gap ID	Aggregated Gaps Description	Anomaly	Development and Testing	Community
A	Greater computational power		X	
B	More efficient and comprehensive communication methods and metrics		X	
C	Better standards for defining roles, responsibilities, and requirements between humans and autonomous systems		X	X
D	S/C need to be able to more quickly detect, diagnose, and respond to anomalies	X	X	
E	Better automated methods and associated hardware for prioritizing and collecting data are needed		X	
F	Autonomous decision-making cannot be done in a way that is trusted to meet performance objective		X	
G	It is difficult to predict the duration and rigor of testing needed to trust autonomous systems		X	
H	It is difficult to gain experience with and trust in operating a more autonomous system in a contested environment		X	
I	It is difficult to build trust in autonomous systems		X	X
J	It is difficult to develop common methods that allow spacecraft to autonomously detect and diagnose anomalies	X		
K	Adoption of new autonomous capabilities by operators takes a long time or does not occur	X		X

Acknowledgments

The authors thank the following for their contributions to and reviews of this paper: Douglas Terrier (NASA), Byron Knight (NRO), Anupa Bajwa (NASA), Shannon Coffey (NRO), Danette Allen (NASA), Fernando Figueroa (NASA), Issa Nesnas (JPL), Kenneth Costello (NASA), Kerianne Hobbs (USSF), Lauren Perry (Aerospace Corp), Lauren Underwood (NASA), Lorraine Fesq (JPL), Mark Harter (MITRE Corp), Merri Sanchez (Aerospace Corp), Richard S. Erwin (USSF), Michele Gaudreault (USSF), and Sherry Olson (NRO).

The authors also acknowledge the support of Alonso Vera (NASA), Benjamin, Seibert (USSF), Brian Thomas (NASA), Christopher Moore (NASA), David Korsmeyer (NASA), Delvin Vannorman (NASA), Dave Anderson (NRO), Jacqueline Lemoigne-Stewart (NASA), Jason Briggs (Linquest), John Day (JPL), Jose Perotti (NASA), Joseph Battle (Tecolote Research Inc.), Joshua Kittle (USSF), Karl Stolleis (USSF), Margret Martin (USSF), Mark Lewis (NASA), Michael Seablom (NASA), Michelle Munk (NASA), Michelle Simon (USSF), Peter Hughes (NASA), Richard Boller (Aerospace Corp), Ronald Clayton (NASA), Sean Phillips (USSF), Terry Fong (NASA), Thomas Pittman (NASA), Thomas Plumb (NASA), and Thomas Niday (USSF).

References

[1] Benjamin, G., Pensado, A., Arney, D., Dempsey, J., Jackson, T., Jefferies, S., Moses, R., Stafford, M., Stillwagen, F., Williams, P., Rodgers, E., Fulton, J., Houghton, N., and Mazarr, A., "Space Science and Technology Partnership Forum: Integration with Commercial In-Space Assembly Activities", AIAA Scitech 2020 Forum, January 6-10, 2020, Orlando, FL.

[2] Save, L. and Feuerberg, B., "Designing Human-Automation Interaction: a new level of Automation Taxonomy", in *Human Factors: a view from an integrative perspective*, Proceedings HFES Europe Chapter Conference, 2012, Toulouse, France.

[3] Ivanco, M. and Jones, C., "Assessing the Science Benefit of Space Mission Concepts in the Formulation Phase", 2020 IEE Aerospace Conference, March 7-14, 2020, Big Sky, Montana.

[4] Jones, C., Ivanco, M., and Deacon, S., "Visualizations to Aid Decision-Making in the ACCP Value Framework", AIAA Scitech 2021 Forum, January 11-15 and 19-21, 2021.