DISSERTATION

INTEGRATING ARTIFICIAL INTELLIGENCE IN HUMAN-RATED SPACECRAFT SYSTEMS FOR LONG DURATION SPACEFLIGHT MISSIONS

Submitted By

Stephanie Anderson

Department of Systems Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, CO

Spring 2026

Doctoral Committee

Advisor: Steven Simske

Susan Bailey

Gregory Marzolf

Marie Vans

ABSTRACT

INTEGRATING ARTIFICIAL INTELLIGENCE IN HUMAN-RATED SPACECRAFT SYSTEMS FOR LONG DURATION SPACEFLIGHT MISSIONS

As human spaceflight missions extend toward deep-space destinations, artificial intelligence (AI) is increasingly suggested as a means to mitigate communication delays, workload, and operational complexity. However, the role of AI in high-stakes, crewed environments remain insufficiently studied from the perspective of end users (crew members). This study reports results from a voluntary survey of a specialized population (N=123), with 25 respondents representing approximately 20% of the target population. Participants included a majority with astronaut/analog astronaut experience and baseline familiarity with AI systems.

Survey results indicate a strong preference for human-AI collaboration over full autonomy, with no respondents favoring fully autonomous decision-making. AI operating as an “Assistant” under human supervision was consistently preferred, though acceptable autonomy levels varied by operational scenario. Respondents placed increasing emphasis on near-perfect reliability as scenario severity increased, strongly valued explainability and transparency, and viewed always-available human override as essential. Trust in AI was moderate rather than absolute, and respondents expressed discomfort with AI making critical decisions without human involvement. Risk tolerance was highly task-dependent, as minimal risk was acceptable for life-critical systems such as life support and navigation,

while greater risk was tolerated for scientific experimentation. Notably, all respondents expressed a desire to participate in the development and training of AI systems they would use operationally.

When compared with existing literature, these findings reinforce emerging consensus that AI for spaceflight should be designed as a bounded, transparent, and human-centered collaborator rather than as an autonomous replacement. The results support the need for graded assurance approaches to AI verification and validation, functional fitness assessments tailored to mission context, and segmentation of the spaceflight landscape by subsystem criticality. Collectively, this work provides grounded guidance for aligning AI system design, assurance, and deployment strategies aligned with operator expectations for future deep-space missions.

ACKNOWLEDGEMENTS

This dissertation represents the end of a long road traveled over many years of determination, resolve, and a thirst for unending knowledge. This post-graduate journey began during a period of personal vulnerability, when returning to structured learning became a way to strengthen my mind and reclaim a sense of direction. I am grateful for the opportunity to transform a challenging chapter into meaningful academic growth, and for the reminder that learning can be both a refuge and a catalyst for healing.

I would first like to convey my appreciation for the support of my advisor, Dr. Steven Simske. From our first encounter to our last he has been a steadfast supporter and motivator for my studies and research, and his constant positivity and excitement has been pivotal in sustaining my motivation and confidence throughout this academic journey. I will be forever grateful for his mentorship.

I would also like to convey appreciation for my doctoral committee members: Susan Bailey, Gregory Marzolf, and Marie Vans. Their feedback and encouragement during my preliminary exam motivated me to expand my thought processes and press forward to the completion of my research.

To my husband, thank you for always being in my corner and encouraging me to pursue my dreams. Your willingness to shoulder countless hours of parenting our young children while I focused on studying, researching, and writing made this dissertation possible.

I have had the distinct honor of having phenomenal managers throughout my professional career, two of which provided me with recommendations for this doctoral program. Thank you to David Gruber (NASA) and David Brueneman (Boeing) for your invaluable mentorship and support in my career. I hope I get the chance to pay it forward in the future and provide such mentorship to others.

To my current employer, Intuitive Machines, and my manager, John Graves, thank you for giving me the space to create, innovate, and explore the boundaries of what is possible on the final frontier. This is a pivotal moment in my career, and I am honored to be surrounded by like-minded dreamers.

I am sincerely thankful to my parents for their constant love and support. Their continuous encouragement throughout my life has given me the strength to push myself academically into realms I never thought I would reach. I am also thankful for my sister, who provides me with endless laughter and reminds me that the ability to relax and enjoy the moment is just as critical as being in performance mode.

Lastly, I extend a heavenly thank you to my grandfather, Luis Esteves. His intense desire to learn, create, and solve problems was passed on to me, and I wish he could be here to share in this academic milestone with me. I know he would be profoundly proud, and I also know he would be asking to borrow my textbooks so he himself could expand his knowledge even further.

DEDICATION

This dissertation is dedicated to all mothers who choose to grow, learn, and persevere for the sake of themselves and their children. May your pursuit of a better future, often carried out quietly, late at night, and alongside countless responsibilities, serve as a testament to the strength of love and the power of determination.

“A computer can never be held accountable, therefore a computer must never make a management decision.” – IBM Training Manual, 1979

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Background

1.1.1 Deep Space Exploration

Deep space human exploration is an important topic that has become increasingly popular over the last decade. NASA has announced plans to return to the moon and from there launch exploration missions into deep space, such as a mission to Mars (NASA, 2023). There are a broad set of human factors challenges that will determine how and when humans will be able to perform their duties during deep space human exploration. This research asserts that many of these challenges can only be addressed by identifying those challenges early in the systems engineering process, i.e. while designing the missions and spacecraft.

Long-duration space missions, such as crewed missions to the moon and Mars, present new challenges for spacecraft design and operation. Communications delays of up to 40 minutes roundtrip between Earth and Mars make real-time human intervention impractical, especially in emergency or time-sensitive situations. As such, spacecraft must possess a high degree of autonomy to handle decision-making, system management, and crew support independent of specialists back on Earth. In addition to on-board expertise, smart computing capabilities on the space vessels are a logical approach to reducing associated risks. Although there are many different aspects of human factors and the effects of deep space travel on the human body that can and should be studied, this research will focus on artificial intelligence can have a positive impact on the impacts of communications delays and decision-making complications potentially seen during deep space travel.

Artificial Intelligence (AI) is emerging as a key enabler of autonomy in everyday life. Amazon's Alexa enables users to access information on-demand while also assisting in every-day tasks such as setting cooking timers and populating grocery lists. Alexa is also capable of learning and becoming more efficient, making it compatible with incremental technological innovations (Fanti et al., 2022). These virtual assistants are not just useful for every-day tasks, but also for medical purposes. A study was conducted in 2023 to assess the effectiveness of using Amazon Alexa as a voice-based format for distributing the Generalized Anxiety Disorder 7 (GAD-7) diagnostic assessment which measures the severity of GAD, finding a positive preference for patients to use this voice-based assessment (Lawson et al., 2024). Another study in 2023 found that using Voice Assisted Technology (VAT) reduced loneliness in older adults, addressing an issue that is comparable to the isolation crew members will face on long-duration spaceflight missions (Jones et al., 2024). Looking towards the future, there are Next-Generation Virtual Personal Assistants (VPAs) being developed that can take several different modes of communication at once, such as voice, gesture, and image recognition (Kepuska et al., 2018). This will allow for the VPAs to assist in medical procedures, vehicle control, counseling, and many other applications (Ibid.).

As proven by these already-developed VPAs, AI systems can process large amounts of data real-time, optimize resource use, perform predictive maintenance, and even support the physical and psychological health of astronauts. This dissertation explores how AI systems can be implemented on spacecraft, focusing on decision-making processes, human interaction, system design, and the role of simulations to validate AI performance.

1.1.2 Historical Significance

The historical trajectory of human spaceflight provides a critical foundation for understanding the emerging role of AI in spacecraft systems. Early programs such as NASA's Project Mercury, which is comprehensively documented in This New Ocean: A History of Project Mercury by Swenson, Grimwood, and Alexander, represent the first large-scale attempts to integrate human decision-making, machine autonomy, and complex system design within the extreme constraints of space. As described in the text, Project Mercury was an "intensive national program mobilizing creative science and technology" to achieve human orbital flight, requiring the coordination of aeronautics, rocketry, and human factors engineering (Swenson et al., 1998). Project Mercury (1958–1963) established the foundational paradigm of spaceflight as a human-machine partnership operating under extreme uncertainty. As Swenson, Grimwood, and Alexander emphasize, the central challenge was not merely achieving flight, but designing a system in which man could function effectively (Swenson et al., 1998). This required careful allocation of responsibility between astronaut, onboard systems, and ground control, which is a problem that remains at the core of spacecraft autonomy today.

1.2 Problem Statement

While extensive research has characterized individual human factors challenges in spaceflight, there remains a critical gap in systematically translating these insights into actionable, integrated system design solutions for deep space missions. Current space system architecture and systems engineering processes insufficiently account for the dynamic interacting nature of human performance degradation in deep space environments. Human factors considerations are often addressed late in the design lifecycle or treated as constraints rather than as primary drivers of system requirements, leading to suboptimal mitigation strategies and increased operational risk. The lived experiences and operational insights of astronauts, who are uniquely positioned to identify performance challenges and system shortcomings, are not consistently or formally leveraged to inform requirements development and design trade-offs.

Based on the current state of the field, there is a need for a systems engineering-driven approach that explicitly incorporates human performance considerations and astronaut input into the early stages of deep space mission design. Addressing this gap will enable the development of actionable, human-centered requirements and design solutions that might mitigate performance degradation, enhance human-system integration, and improve the chance of mission success for future deep space exploration missions.

1.3 Research Objectives

This dissertation seeks to uncover how artificial intelligence may be capable of bridging the gap between the impact of stressors on the human body during long duration spaceflight and the critical skills required to operate a spacecraft in deep space. The research objectives are:

1. To explore how human performance is degraded in deep space and what special considerations need to be made for that form of space travel.
2. To address how systems engineering methods can be used to address human factors issues for deep space travel.
3. To leverage AI and astronaut input to drive actionable requirements to solve human factors and systems engineering problems for deep space travel and implement those requirements in system design.

1.4 Significance

This research, along with the NASA study and results, will make a great contribution to the current research on how to make deep space missions successful. It will add data from crew member feedback, which is something that appears to be lacking in the current state of the field. There is some medical data available from the crew members, such as from the Twin Study which will be further reviewed in the literature review section, but there is limited information on the actual thoughts and feelings from the crew members themselves. The goal is to find the areas most relevant where we can use astronaut input to drive actionable requirements to solve human factors and systems engineering problems for deep space travel while also leveraging AI solutions. Future research can then be built upon the findings from this study to better prepare those who are in the developmental stages of designing hardware and missions to deep space.

1.5 Dissertation Structure

This dissertation provides insight into how crew members training for and flying deep space missions want AI to be involved in their spacecraft system design. To address this issue, the content of this dissertation is structured as follows:

Chapter 1: Introduction provides background on the topic, the problem statement, research objectives to help guide the study, and the significance of the research. This chapter serves as the foundation to explain why there is a need to research how astronauts want AI to be involved in their spacecraft system design and how AI should be implemented operationally.

Chapter 2: Literature Review investigates space stressors and psychology levied on crew members, human decision-making during spaceflight, human factors and performance, human-systems interactions, and the potential value of AI in space systems.

Chapter 3: Theoretical Framework explores how AI is a systems element, focusing on design and simulations. This chapter also gives a comprehensive comparative analysis of current AI systems in spaceflight, as well as key areas where to avoid role reliance on AI.

Chapter 4: Methodology introduces the research design, experimental analysis, theoretical exploration, analytical approach, ethical considerations, and limitations of the study.

Chapter 5: Experiment and Results introduces the experiment and lays out the results of the survey, identifies the interesting and notable trends, insights from qualitative feedback, and what limitations and challenges were identified.

Chapter 6: Analysis and Disc-ussion outlines the interpretation of the key findings, comparisons of the findings with existing literature, and what implications the results may have on verification and validation of AI, AI functional fitness assessments, and segmentation of spaceflight landscapes for AI applications. This chapter ends with recommendations for future research.

Chapter 7: Conclusions summarizes findings of the study and reinforces areas for future research. The chapter emphasizes what challenges should be addressed before AI is implemented in human rated spacecraft for long duration missions.

CHAPTER 2: LITERATURE REVIEW

2.1 Space Stressors and Psychology

2.1.1 Introduction

Preparing humans for deep space travel requires more than technological innovation; it demands a high understanding of the psychological and behavioral challenges astronauts may face during months or years far from Earth orbit. NASA categorizes behavioral health as one of the top risks for exploration missions, recognizing that human performance is directly linked to mission success (NASA Human Research Program, 2023).

Astronauts will face extreme stressors during long duration spaceflights which will present a unique challenge. These stressors include "radiation, gravity fields, hostile/closed environments, isolation and confinement, and distance from Earth" (Landon et al., 2024). These stressors will have a significant impact on both physiological and psychological functioning, affecting not only personal performance but team performance as well. In a study conducted in mice, female mice irradiated with 50 cGy resulted in "impaired object recognition", and the combination of social isolation and HU (hindlimb unloading, which simulates microgravity) housing stressors resulted in the same conclusion (Rienecker et al., 2023).

2.1.2 Isolation, Communications Delays, and Autonomy

One of the most significant stressors that will be unique to deep space exploration is prolonged isolation from Earth. For example, during a mission to Mars astronauts may be up to 400 million kilometers away, resulting in two-way communication delays of up to 40 minutes. This delay fundamentally shifts the dynamics of teamwork, troubleshooting, and emotional support that veteran astronauts are accustomed to having while in low Earth orbit (LEO). Crews will need to be capable of operating with greater autonomy, making critical decisions without immediate input from mission control. NASA's Human

Exploration Research Analog (HERA) studies show that isolation, combined with operational pressure, increases risks of mood disruption, cognitive fatigue, and interpersonal conflict (Slack et al. 2016).

Communication delay can also alter team dynamics. Crew-ground interactions on the International Space Station (ISS) serve as both technical support and social stabilization, acting as a buffer against tension and stress. In deep space, this buffer weakens significantly. As Kanas and Manzey (2008) note, the absence of immediate communication increases feelings of detachment and requires crews to rely more heavily on internal cohesion and problem-solving capabilities. Therefore, deep space missions must incorporate psychological training emphasizing autonomy, conflict management, and shared mental models to compensate for reduced real-time support, as well as implement AI solutions to aid in psychological management.

2.1.3 Confinement and Sensory Monotony

Spacecraft used for deep space travel will be compact and resource efficient. While functional, such environments impose sensory monotony which is a mental and emotional state of fatigue caused by a lack of varied or stimulating sensory input. Research from Antarctic overwintering expeditions, which are commonly used as analogs for space, shows that monotony induces cognitive slowing, irritability, and decreases motivation (Palinkas & Suedfeld, 2021). HI-SEAS analog studies further demonstrate that restrictive habitats can trigger boredom, stress-related fatigue, and interpersonal strain, particularly after the initial novelty wears off (Rosenberg et al., 2022).

Sensory deprivation carries operational risks as well. When environments lack variation, individuals may experience attentional lapses or over-focus on irrelevant details, both of which are dangerous during complex spacecraft operations. NASA is addressing this through habitat designs featuring adjustable lighting, digital windows, personalized workstation environments, and scheduled environmental changes (NASA Behavioral Health & Performance, 2022). These interventions aim to preserve cognitive engagement and emotional well-being over long durations.

2.1.4 Workload, Fatigue, and Circadian Disruption

Sleep and circadian disruption are persistent issues in spaceflight and are expected to intensify in deep space missions. In LEO, astronauts sleep on average six hours per night, significantly below recommended levels (Barger et al., 2016). Without natural solar cycles and with mission demands dictating irregular schedules, circadian misalignment becomes common. Disrupted circadian rhythms impair cognitive performance, emotional regulation, memory, and decision-making, which are all critical capabilities in a high-risk space environment.

Deep space vehicles present additional challenges, specifically due to the lighting, hardware noise, and temperature fluctuations. NASA's introduction of the Solid-State Lighting Assembly (SSLA) aims to regulate melatonin production through carefully designed light wavelengths, supporting healthier circadian alignment (NASA Human Research Roadmap, 2022). Despite these improvements, missions to Mars will require even more sophisticated countermeasures, including real-time fatigue monitoring which can be enhanced with the implementation of AI solutions.

2.1.5 Team Cohesion and Interpersonal Dynamics

Effective teamwork is arguably the most critical psychological factor in deep space missions. Long duration confinement heightens emotional sensitivity and increases the likelihood of interpersonal conflict. NASA's Behavioral Health & Performance team has documented that crew cohesion, compatibility, and communication style are strong predictors of mission performance (NASA BHP, 2022). Mars500 data reveal that interpersonal tensions typically peak during the mid-mission "third-quarter phenomenon", when motivation dips and irritability increases (Sandal, 2001).

To prepare for these dynamics, crews undergo psychological training, conflict resolution exercises, and cultural competency development. Additionally, mission planners emphasize balanced team composition with diversity in expertise, communication style, and temperament to help maintain group adaptability and reduce friction.

2.1.6 Space Fog

Another phenomenon that has been discovered to negatively impact astronauts in space is "space fog". Space fog is a condition which affects astronauts, typically first-time-flyers and those who have recently arrived in microgravity, and produces symptoms which include "befuddlement, altered time sense, dizziness, disorientation, mental slowing, poor concentration, a sense of de-realization, and the press to perform tasks very slowly and precisely” (Welch et al., 2009). Although this condition is typically experienced in the first few days of spaceflight until acclimation occurs, it can also be experienced at any time during the mission. A study conducted by Welch et al. demonstrated microgravity slows visual-motor behavior, contributing to the symptoms of space fog.

While it is critical for us to understand how behavior and the human brain are affected by long duration spaceflights, and microgravity/zero-gravity in general, it is equally crucial to develop mitigations and resiliencies to combat these effects to assure crew safety and mission success.

2.2 Decision-Making during Spaceflight

The increase in communication delay and forced autonomy during long duration missions will create difficulties with working in sync with mission control for real-time decision making. Spaceflight involves complex and often time-critical decisions, particularly during long-duration missions. Astronauts exploring deep space will need to have adequate training to be able to handle medical emergencies and spacecraft hardware/software emergencies on their own without immediate communication with ground teams in mission control. Medical emergencies, as well as many other emergencies astronauts may encounter while in deep space, require quick and effective decision-making. If a decision to change your mind about something is made within 100 milliseconds of being presented with the choice, the change of mind will succeed in altering the original course of action. However, if it takes at least (or more than) 200 milliseconds, the chances of the change succeeding are significantly less (Beres & Bell, 2022). Based on that research, there is no room for dwelling on making a decision. When it comes to making simple decisions, on average older adults take longer to process information and make decisions than younger adults. In complex decision-making scenarios, age is associated with faster decision making (Löckenhoff, 2011). Although there has been a great deal of research done into the measurement of human decision-making, this kind of research has not been conducted and applied to include the stresses of deep space travel.

One saving grace in the human brain is neuroplasticity. Neuroplasticity is the brain's ability to reorganize its structure as a result of trauma, new experiences, or new information. This brain functioned was observed during the NASA Twin Study, where "overall their data show plasticity and resilience for many core genetic, epigenetic, transcriptional, cellular, and biological functions" (Garrett-Bakelman et al., 2019). Neuroplasticity is a natural form of resiliency in humans and can contribute to naturally combating the trauma spaceflight inflicts on the brain, however it is not capable of being the sole mitigation factor.

To assist with the degradation of decision-making capabilities, autonomous systems must be capable of real-time decision-making in a variety of scenarios. In fact, a significant portion of research into how decision-making difficulties will be mitigated during long duration spaceflight assume astronauts will be aided by AI, for example in the areas of "construction, exploration, mining, and mapping" (Rappaport et al., 2022). These key areas/scenarios include:

- Navigation. AI algorithms can adjust a spacecraft's trajectory, navigate around possible conjunctions, and plan fuel and time efficient routes based on available data.
- Resource management. AI systems can monitor and optimize the use of limited resources such as fuel, water, and oxygen, ensuring the sustainability of a long-term mission.
- Emergency responses. In the event of hardware or equipment failure, AI can diagnose the problem, suggest solutions, or autonomously take corrective action. This can also be applied in a health emergency, where AI can act as a virtual medical professional.

Given the complexity of long-duration space missions, AI decision-making systems must be robust, reliable, and adaptable to unforeseen circumstances. The robustness, reliability, and adaptability will be quantified during the fault injection portion of system testing to observe how resilient and accurate the AI system is. The crew members must be able to rely on, without hesitation, the actions the automated system would take. The reliability factor can only be achieved with a robust integrated design along with vigorous simulations to test functionality and reliability.

2.3 Human Factors and Performance

2.3.1. Effects of Spaceflight on Human Performance

Deep space missions will produce a spectrum of cognitive overload and underload. The mission may produce short periods of overload, including high periods of stress, fatigue, and cognitive demand, followed by potentially long periods of underload, which can cause depression and loneliness (Holden et al., 2021). This overload of swinging emotional states can have a negative impact on human performance. As it stands with current human space travel to LEO, astronauts have near-immediate access to communications with mission control which therefore affords them access to psychological assistance, if needed, as well as social interaction. Given the increased communications delays the further missions extend past Earth, astronauts will need to be enabled to perform their work autonomously and without the constant potential access of mission control. For astronauts to be capable of this level of autonomy, developing tools to assist with task execution, problem solving, and behavioral health will be paramount. These tools will need to "rely on more automation, advanced computational techniques, and integrating varied and various data sources" to "support crew autonomy and a new approach to the entire multi-team system" (Marquez et al., 2023). Human factors must remain at the forefront of this technological development to ensure humans are kept in the loop and are afforded consideration in design and implementation.

2.3.2. Performance Shaping Factors (PSF)

Performance Shaping Factors (PSF) are conditions in an individual's environment and internal state that influence their performance. These can be organizational, environmental, or related to lifestyle, and can significantly impact human performance. Through NASA's effort to quantify human spaceflight performance in the crew health and performance system, PSFs that have been identified are exercise, sleep, and hydration (Matar et al., 2024). Additional PSFs were identified while analyzing the importance of human reliability in human space flight, which include adequacy of organization, working conditions, adequacy of operational support, availability of procedures/plans, number of simulations goals, available time, adequacy of training and experience, and crew collaboration quality (Hamlin, 2010).

Taking these identified PSFs into consideration, an evaluation must be made on the need to reshape based on the differences long duration spaceflights pose compared to missions to LEO. This entails reshaping the PSFs themselves or reshaping how they are mitigated in the deep space environment. Below is a table of some generalized suggested PSFs for astronauts traveling on deep space missions, how they may be mitigated, and if mitigations can be assisted by AI:

Table 1. Performance Shaping Factors (PSFs), Mitigations, and AI Involvement

PSF	Mitigation	Can be assisted by AI?
Fatigue/Workload	Autonomy-optimized scheduling and workload balancing	Yes
Adequate training	Onboard continuous learning	Yes
Radiation	Advanced shielding	No
Autonomy level	New tool/technology development	Yes
Habitat size/layout	Mindful design, human factors involvement	No
Team cohesion/interpersonal dynamics	Astronauts train and fly together	Yes
Skill decay	Onboard continuous learning	Yes
Resource availability/scarcity	Tools for optimizing resource use	Yes

While the above table is not intended to be a complete, comprehensive list of PSFs, it provides insight into high level PSFs and conveys how AI can be an influential component to providing mitigations.

2.4 Human Systems Integration

When considering spacecraft system design and how that design will evolve as the system is developed for deep space missions, keeping the human in the loop will be paramount to ensure crew performance is being optimized. It will include understanding the overall role of humans within the technical settings keeping the "interaction between humans and technology as a system" (Lai & Selby, 2016).

Human-in-the-loop (HITL) involvement for spacecraft system design is essential for deep space missions because human performance, limitations, and adaptability ultimately shape mission success. As Lai and Shelby (2016) emphasize, even the most technologically advanced systems cannot replace the nuanced judgement, improvision, and problem-solving abilities that humans bring to unexpected situations. Deep space missions introduce unprecedented levels of isolation, communications delays, and operational uncertainty. Under these conditions, the design of spacecraft must prioritize human factors from the earliest engineering phases, not just as an afterthought during HITL testing post-development. Without meaningful human integration, systems run the risk of being misaligned with crew needs, too complex to operate safely, or unusable during high-stress situations when clarity and ergonomics matter the most.

Another critical aspect is error mitigation. Human error in spaceflight often emerges from poorly designed interfaces, excessive workload, or confusing automated behaviors. Effective human-systems integration (HSI) reduces these risks by ensuring that displays, controls, and automation are intuitive, consistent, and supportive rather than overwhelming. In deep space environments, crews will not have real-time support from Earth due to potentially lengthy communication delays. That means spacecraft systems must be designed so that astronauts can autonomously detect, diagnose, and recover from anomalies. Keeping humans in the loop by being actively involved in spacecraft monitoring and decision-making ensures that their cognitive strengths are leveraged rather than sidelined by automation. The introduction of automation or AI tools should be added as an assistant, not as a replacement for human knowledge.

Long duration missions demand systems that accommodate human variability over time. Fatigue, stress, and adaptation to altered gravity can change how astronauts perform tasks or interpret information. HITL design makes spacecraft resilient to these fluctuations by incorporating flexible controls, adaptable interfaces, and systems that evolve with crew needs. It also ensures that operations remain transparent, as astronauts must understand what automated systems are doing and why. When people trust and comprehend their tools, they are more capable of intervening effectively, especially when in the face of novel or cascading failures far from Earth.

Ultimately, Lai and Shelby (2016) reinforce that deep space exploration is not solely a technological endeavor – it is a human one. Keeping humans in the loop ensures that spacecraft are not only high-performing machines but also safe, usable, and supportive environments for the people who must live and work within them for potentially years at a time. This integration is fundamental to mission reliability, safety, and the long-term success of human exploration beyond Earth orbit.

2.5 Value of Artificial Intelligence

2.2.1 Enhancing Autonomy

The primary value of AI in spaceflight is to increase spacecraft autonomy. AI systems can process large amounts of data gathered real-time from sensors and other instruments and make decisions that would otherwise require human intervention, which would result in a greater time between the “event” and the “response”. For example, if one of four navigation sensors takes a hit from foreign object debris (FOD) and begins to have false readings, an AI system would be able to instantly detect the problem, remove that sensor from the navigation calculations, rely solely on the functioning navigation sensors, begin diagnostics on the malfunctioning sensor, and notify ground teams back on Earth of the problem. If an AI system was not available, the navigation system would have to wait until a crew member was alerted to the diverging data from the malfunctioning sensor, assess the data, take corrective action, and notify ground teams. Not only would this result in a delay of corrective action, but it would take the crew member away from another potentially important task that they may have been working on. This autonomy would also reduce the reliance on the Earth-based mission control, allowing the spacecraft to function more independently over extended periods of time.

2.2.2 Optimizing Resource Utilization

AI can optimize the allocation and consumption of critical resources, such as fuel, water, and oxygen. Predictive models can forecast resource use under various mission conditions, allowing for more efficient planning and the prevention of resource shortages during long-duration missions. While oxygen and water are continuously generated using an electrolytic oxygen generator (splitting water molecules into hydrogen and oxygen) and a water processor assembly (taking wastewater and turning it into drinking water), if either of these pieces of hardware malfunctions and the ration of oxygen or water is necessary, an AI system could quickly determine a rationing schedule to keep the crew safe and healthy.

2.2.3 Supporting Crew Well-being

Long-duration missions put excess stress on the human body and expose astronauts to various physical and mental health risks. AI can monitor the crew's health, identify patterns that may indicate stress or illness, and provide recommendations for maintaining well-being. This can be provided on an individualized, or "custom medicine", basis (Johnson et al., 2021). AI can also be used to adapt this "customization" over time; e.g., as the effects of space travel begin to be felt on astronaut physiological systems such as the eyes, the musculoskeletal system, and the cardiovascular system, among others. Perhaps most crucially, AI can be used to address changes in affect and mental health. As previously described for alleviating the feeling of loneliness, AI-driven mental health support systems can engage astronauts in dialogue, assess their cognitive and emotional state, and offer therapeutic interventions when necessary, leading to an overall healthier crew.

2.2.4 Powering Human Exploration

Human exploration of deep space requires technological systems capable of operating far from Earth for extended durations, often without real-time support. Among the most critical challenges is the development of reliable, autonomous, and sustainable power systems. Energy enables every aspect of human activity in space, from life support and propulsion to communications and scientific operations. Deep space missions to the Moon, Mars, and beyond demand new approaches to energy generation, storage, distribution, and intelligent management. The literature reveals that next-generation AI-enabled power systems, advanced batteries, solar propulsion technologies, and autonomous spacecraft control are essential components of future exploration architectures (Frank, 2019; Li & Li, 2011; Oche et al., 2024; Straub 2011). As human missions venture father into the solar system, power systems must become more resilient, autonomous, and deeply integrated with spacecraft operations.

2.2.4.1 AI-Enabled Power Management and Autonomy

As missions extend farther from Earth, communications delays make real-time ground monitoring and control of power systems impossible. AI and autonomous control systems have become essential to ensure safety and efficiency. Frank (2019) highlights the need for onboard intelligence that can plan energy usage, diagnose faults, reconfigure systems, and execute repairs without ground intervention. These capabilities reduce astronaut workload and increase mission resilience.

Straub (2011) provides a comprehensive review of spacecraft AI control systems, noting their expanding use in mission planning, fault detection, attitude control, and resource allocation. AI systems allow spacecraft to anticipate power peaks, distribute energy more efficiently, and respond to anomalies before they escalate into failures. In crewed missions, this autonomy supports life support systems, environmental controls, propulsion, and habitat utilities, making the power subsystem an active participant in mission safety rather than a passive component.

The role of AI also extends to habitat infrastructure. Intelligent power networks can balance the needs of life support, navigation, communications, scientific instruments, and surface operations. AI-driven microgrids, as described by Frank (2019), ensure uninterrupted energy flow and enable dynamic load-shedding during emergencies. As deep space bases become more complex, incorporating surface rovers, scientific laboratories, and in-situ resource utilization (ISRU), AI take on an even greater role in managing interconnected power demands.

2.2.4.2 Systems Integration and Mission Architecture

Power systems in deep space are not isolated subsystems; they are integrated into the overall mission architecture. Li and Li (2011) argue that deep space power design must consider mission phases holistically, taking into consideration launch, transit, landing, surface operations, and return. Different mission phases require different energy requirements and having an AI tool that can organize and plan power resources would be a large asset to any deep space mission.

Oche et al. (2024) also note that spacecraft power systems must be resilient during deep space environmental events such as dust storms on Mars, solar radiation events, and wide thermal fluctuations. Robust architecture ensures that even when environmental conditions limit solar energy, critical systems remain powered through stored energy or alternative generation sources.

2.2.5 Crew Resilience

2.2.5.1 Reducing Crew Workload and Supporting Cognitive Resilience

One of the most significant ways AI will enhance crew resilience is through workload reduction. Deep space crews must manage life-support systems, navigation, resource utilization, power management, scientific operations, and maintenance, all of which in an unforgiving environment. Prolonged cognitive overload can contribute to stress, fatigue, and performance degradation. Frank (2019) emphasizes that future mission architectures require AI systems capable of autonomously planning, monitoring, and executing mission activities. These intelligent systems can schedule tasks, allocate power resources, prioritize scientific operations, and even troubleshoot anomalies without constant crew intervention. By reducing the cognitive burden on astronauts, AI enables them to maintain clarity, focus, and psychological well-being.

AI-enabled decision support also plays a key role in cognitive resilience. Straub (2011) describes how onboard AI can synthesize spacecraft data, assess trends, and recommend courses of action. This reduces the mental strain associated with interpreting large, complex data streams across spacecraft systems. In high-stress situations, such as equipment failures or hazardous environmental events, AI's ability to provide clear, timely recommendations allows crews to respond quickly and effectively without becoming overwhelmed. This type of cognitive framework is especially important in deep space, where communication delays prevent immediate guidance from Earth.

2.2.5.2 Autonomous Fault Management and Safety Assurance

Crew resilience is deeply tied to perceived safety and environmental stability. When astronauts trust that their systems can autonomously identify and resolve faults, their stress levels decrease, allowing them to focus on mission-critical tasks. Across the literature, autonomous fault detection, isolation, and recovery (FDIR) emerges as a cornerstone of resilient mission operations (Frank, 2019; Straub, 2011). AI-driven monitoring systems can predict failures before they occur, reroute power, adjust environmental controls, or reconfigure spacecraft subsystems without requiring human intervention. Oche et al. (2024) highlight the integration of AI into power systems to ensure uninterrupted energy availability—one of the most psychologically important factors for crew safety and stability. Maintaining a reliable power supply contributes directly to psychological resilience by ensuring that habitat systems, lighting, thermal control, communications, and life support remain stable even during anomalies.

Improved safety through AI autonomy also helps mitigate stress from isolation. In deep space, astronauts cannot rely on immediate support from mission control. Knowing that the spacecraft can handle emergencies without real-time monitoring enhances crew confidence and reduces anxiety. AI systems effectively serve as a “virtual expert team” embedded within the spacecraft, reinforcing the crew's sense of security and operational stability.

2.2.5.3 Supporting Psychological and Social Resilience Through Intelligent Habitats

As surface habitats become more complex, AI will be deeply embedded in environmental controls, scheduling, and social support structures. Intelligent habitats will adapt lighting, temperature, and noise levels to improve sleep, alertness, and mood, which are critical factors in long-duration confinement. While the articles focus on AI control systems and power management, the principles extend naturally to habitat autonomy. Li and Li (2011) explain that spacecraft control systems are increasingly integrating AI to manage environmental complexity, providing more stable and adaptive living conditions. Such systems can detect early signs of environmental stress such as temperature fluctuations, rising carbon dioxide, or insufficient humidity, and adjust conditions automatically, ensuring crew comfort and psychological well-being.

AI can also mediate social and interpersonal dynamics. Long-duration missions often experience interpersonal tension, communication fatigue, and periods of isolation. AI-driven communication monitoring, scheduling algorithms, and adaptive crew-support systems can mitigate conflict by ensuring balanced workloads, synchronized rest periods, and optimized group task assignments. Straub’s (2011) analysis highlights the growing capability of AI to evaluate system states, predict trends, and adapt behavior dynamically, which are skills that apply well to social systems within a spacecraft.

CHAPTER 3: THEORETICAL FRAMEWORK

3.1 Introduction

This chapter establishes the theoretical framework which will act as the foundation of the research for implementing AI in spacecraft for long duration missions. The design, human-system interaction, manual override, and simulation steps will establish the framework used to design the survey questions in the research conducted for this study. It is also important to acknowledge the current usage of AI in space, as well as identify where to avoid sole reliance on AI.

3.2 AI as a System Element

3.2.1 Design

Designing an AI system for spaceflight involves creating algorithms that can perform real-time analysis under resource constraints, such as limited computational power, limited energy, and a bounded knowledge base. Due to the physical distance between spacecraft on long-duration missions and Earth, access to internet may not be instantaneous. Therefore, the information available to an AI system would have to be pre-programmed and stored onboard the spacecraft. This knowledge base would need to be robust, encompassing all of the information needed in order to adequately perform the intended tasks. Updates can be periodically made to the knowledge base when a connection with Earth can be established, such as via the orbiters circling Mars that can act as relay stations in the interplanetary system (NASA's Mars Odyssey, MAVEN and Mars

Reconnaissance Orbiter, and European Space Agency's Mars Express and Trace Gas Orbiter) (Platt, 2024). However, once an "interplanetary internet" is created the AI system may be able to access information from the internet real-time, or at the very least intermittently, for more frequent updates (Burleigh et al., 2002). Other key design considerations include:

- Fault tolerance. The spacecraft AI must be able to handle hardware or software failures, recover autonomously, and have access to key spare hardware.
- Data processing. AI systems need to have the ability to process data efficiently real-time, using edge computing to reduce latency.
- Robust algorithms. AI models should be capable of handling diverse and unpredictable data, such as environmental changes or system anomalies.

3.2.1.1 Human and System Interaction

AI systems on spacecraft must interact seamlessly with human crew members. Interfaces that enable astronauts to understand AI recommendations, and where those recommendations stem from, are essential to avoid confusion or lack of trust. The design should follow Nielsen's 10 Usability Heuristics for design interfaces: visibility of system status, match between system and the real world, user control and freedom, consistency and standards, error prevention, recognition rather than recall, flexibility and efficiency of use, aesthetic and minimalist design, help users recover from errors, and access to help and documentation (Nielsen, 2024). As such, the AI system should continuously provide feedback on spacecraft status, including navigation, life support, and communications. This should include real-time updates on trajectory, fuel consumption levels, oxygen levels, and any detected anomalous conditions. Terminology and symbols familiar to the crew members should be used, consistent with labels and standard spacecraft terminology.

3.2.1.2 Manual Override

Despite the desire to have autonomy provided by the AI system, there must always be a mechanism for manual override. There may be situations where human judgement is required, and astronauts need the ability to bypass the recommendation by the AI system and have direct control of the spacecraft. This can be in situations where the “human” aspect must be considered, or in situations where the AI system may be malfunctioning. This will ensure the crew retains ultimate authority over mission decisions. Manual override may also be necessary when conducting maintenance on the AI system, or while the AI system is undergoing a software update. Currently, there is an ongoing study at Johnson Space Center to understand the impact of spaceflight on a crew’s ability to perform manual override tasks, and to examine how adaptive vestibular and cognitive function relate to changes in manual crew override proficiency in a simulated lunar landing (Bollinger et al., 2024). The need for this and similar studies proves the expectation that crew members may need to perform manual overrides in various situations and provides data on how crew members can mitigate risks introduced by human error.

3.2.1.3 Data, Security, and Operations

AI systems are energy intensive and require a lot of power, with the demand growing as system intensity grows. Power requirements stem from data storage, computational power, processing power, displays, and operations. These challenges are similar to the challenges faced with deployment of the Internet of Things (IoT), depending on large amounts of labeled training data to produce meaningful output (Goodwill et al., 2023). Due to these challenges, in recent years there have been demonstrations of spacecraft AI functionality with relatively simple algorithms due to limitations of onboard processing capability (Goodwill et al., 2023). New technologies in energy storage systems should be explored in order to accommodate the extra load requirements, such as hybrid energy storage systems (Marín-Coca et al., 2024).

The AI system should have a concept of operations (CONOPS) that includes security operations (SECOPS) and data operations (DECOPS). SECOPS will need to include mitigations for cybersecurity vulnerabilities, those identified as exclusively an AI vulnerability and those identified as fundamental software infrastructure issues (GOV.UK, 2024). Mitigations will need to be put in place beginning at the design phase and going through the development, testing, deployment, and maintenances phases, while allowing for the AI system to continue to “grow” and “learn” during mission operations. Examples include adequate threat modeling, robust security architecture, data privacy safeguards, strong access controls, solid infrastructure, encryption, and insider threat monitoring (GOV.UK, 2024).

3.2.2 Simulation

Through the efforts of rigorous simulation, trustworthiness will be developed between the AI and the user. This will reinforce the user’s ability to trust the answer an AI gives to a question the user does not know the answer to, which is one of the fundamental issues with using an AI system: believing the AI is correct when the user does not know the answer themselves.

3.2.2.1 Parameters

Simulating AI systems before deployment is essential to ensure robustness and reliability for long-term missions in space. Simulations must cover a wide range of parameters, including:

- Mission duration. Long-duration missions require simulations that can test the AI's performance over extended periods of time, both with and without crew interactions.
- Environmental conditions. The AI must be tested in a variety of space environments, including deep space and planetary surfaces with variable gravitational conditions and radiation levels.
- System failures. Simulations should introduce system anomalies to test the AI's ability to respond to the unexpected. These simulations should feature a Monte Carlo method, where a broad set of algorithms relying on random sampling are used to create, and subsequently solve, problems, predicting all of the ranges of inputs (Rößler et al., 2022).

3.2.2.2 Venues

Simulations for spacecraft AI systems should be conducted in various conditions/venues, such as:

- Ground-based simulations. Earth-based testbeds, such as analog missions (e.g., Human Exploration Research Analog, HERA), can simulate long-duration mission conditions to test both the AI system itself and the human interaction.
- Virtual simulations. Other ground-based simulations can be conducted without user interaction, having the AI system run its own simulation based on a set of parameters such as various mission objectives and navigation requirements. Advanced flight software environments can replicate spaceflight dynamics and provide a controlled way to test AI algorithms and their decision-making capabilities.
- LEO missions. ISS has long served as a testbed for science, both physical and virtual/software. The ISS is an ideal venue as it provides many actual space conditions (significantly reduced gravity, etc.) with the caveat of a lack of communications delays and controlled radiation exposure.

3.2.2.3 Post-Deployment

The simulations should not solely occur during pre-deployment testing. Since the AI system will be continuously learning, growing, and implementing software updates, periodic self-checks will be required to assess continuing functionality. A set of pre-determined test parameters can be selected to test the AI system incrementally, such as once a week, once a month, after every software update, etc., depending on the recommendations of the development team.

3.3 Comparative Analysis of Current AI Systems in Spaceflight

Several AI systems have been developed for space applications, with varying levels of functionality and autonomy. A comparative analysis of these relevant systems reveals the strengths, weaknesses, and challenges of different designs.

NASA's Autonomous Systems and Operations (ASO) program is focused on integrating AI into spacecraft operations. ASO AI systems assist in real-time problem solving and provide operational support to astronauts. Their strength lies in decision-making support and fault detection but still require significant human oversight. This project also reveals the importance of multiple venues of simulation, as it has been tested on the ISS in 2014-2016, in the Integrated Power, Avionics and Software (IPAS) facility at Johnson Space Center, and in a testing facility during Exploration Flight Test 1 (EFT-1) (NASA, 2021).

IBM's Crew Interactive Mobile Companion (CIMON) was developed and launched in 2018 to the ISS as an AI-powered assistant. While CIMON provides crew support through voice interaction, its limitations include reliance on pre-programmed responses and limited autonomy in decision-making. It was a quickly completed project, only taking about two years to develop, and therefore IBM did not install a complete cloud service on the ISS and it instead relied on the ground IBM cloud and satellite connection to enable the CIMON software architecture and end-to-end connectivity (Eisenberg, 2024). In 2019, CIMON-2 was launched to the ISS to demonstrate an updated version of the AI system, testing autonomous flight capabilities, voice-controlled navigation, and the ability to understand

and complete tasks (Airbus, 2020). This project aimed to research whether AI assistants, such as CIMON-2, can be used to reduce crew member stress.

3.4 Where to Avoid Sole Reliance on AI

AI provides an attractive solution for automating decision-making in the high stress environment of long duration space missions. However, there is a question of where the line should be drawn between automation and the necessity of human interactions for complex decision-making. Along with the integration of manual override, there is a need for a definition of where automation should be avoided within the system.

There are stages of flight where critical decision-making requires human oversight. Launch and landing decisions, along with collision avoidance maneuvers, are examples of situations where human oversight is crucial to ensure that nuanced ethical considerations, intuition, and unforeseen edge cases are handled appropriately. For example, if the spacecraft is headed on a collision course with space debris and the automation calculates a debris avoidance maneuver that will put the spacecraft into a trajectory increasing the travel time by several months, that would not be something the crew members would be likely to choose due to the mission length extension. Instead, the automated system could present the crew members with several options with a risk-reward analysis and allow the crew members to make the decision based on what they deem is most appropriate. It may not be as simple as "pick the trajectory with the shortest added time", since the additional time may result in necessary mission abortion, collateral risk, and other outcomes perhaps unanticipated by the AI. Therefore, human intuition and anticipation of situations on which the AI has not been trained are also factors to consider. AI also lacks moral reasoning, being a purely intelligence-based system. An automated system may require human interaction when deciding the distribution of critical resources, such as oxygen, food, and water, and may value the overall mission objectives above the health of individual crew members.

There is also the danger of over-automation thereby causing the crew members to rely on the AI excessively, causing an expense of crew skills and potential complacency or even resentment. In addition, excessive reliance on AI may erode the crew members' problem-solving and critical-thinking skills that are highly needed to retain manual control and procedural knowledge in case of emergencies when AI might become damaged or fail.

CHAPTER 4: METHODOLOGY

4.1 Introduction

This study aims to assess the usability, trust, and risk perceptions associated with AI integration in spacecraft for long-duration missions. The research seeks to understand the preferences, experiences, and expectations of individuals with varying backgrounds in spaceflight and AI, providing insights into optimal AI roles, reliability standards, and human-AI collaboration in both low and high stakes environments.

4.2 Research Design

The following defines the research hypothesis and null hypothesis:

Hypothesis: In the condition of deep space travel, astronauts would prefer the override function to be human as opposed to artificial intelligence, and astronauts would also prefer for emergency responses to be taken by a human rather than artificial intelligence.

Null hypothesis (H_0): There is no significant difference in astronaut preference between a human-controlled override or emergency response and an AI-controlled override or emergency response.

The following research questions (RQ) aim to be addressed through this research:

RQ1: What are the human performance differences between travel to low Earth orbit and deep space, and what special considerations need to be made for deep space?

RQ2: How can integrating artificial intelligence be used to address human factors issues for deep space travel?

RQ3: Can we use astronaut input to determine if the use of artificial intelligence would be a reliable and dependable solution to solve human factors and systems engineering problems for deep space travel?

RQ1 and RQ2 are addressed through the literature review. To test the hypothesis and explore RQ3, a cross-sectional, web-based survey was developed and distributed to participants with experience in training for space missions or participating in missions to LEO. The survey comprises 34 total questions which include multiple choice and open-ended responses. The questions are organized into thematic sections: participant background, AI familiarity, scenario-based decision-making, risk tolerance, trust, and attitudes toward AI development. The survey was accessible via a unique web link, ensuring voluntary and anonymous participation. The survey was designed to capture both quantitative and qualitative data, allowing for a comprehensive analysis of attitudes, experiences, and recommendations.

4.3 Experimental Analysis

The survey collected responses from 27 participants between June 11, 2025, and July 29, 2025. Respondents included analog astronauts and astronauts with and without spaceflight experience. Key variables measured include experience level, AI familiarity, preferred AI roles (collaborative, human-driven, fully autonomous), required reliability and transparency, human override preferences, confidence in choices, trust in AI, comfort with AI autonomy, and risk acceptance in various mission-critical scenarios. Scenario-based questions were repeated across different mission contexts to assess consistency and variability in responses. Open-ended questions solicited additional comments on AI integration, with responses subjected to thematic and sentiment analysis.

4.4 Theoretical Exploration

The survey was grounded in established theories of human-AI interaction, trust in automation, and risk management. The design drew on literature regarding how AI can be a helpful tool, especially for decision-making, during long-duration missions and the importance of a trustworthy human-AI interaction, specifically when crew members may be in a state of psychological stress in high-stakes environments. Scenario-based questions were crafted to elicit responses aligned with these theoretical frameworks, enabling exploration of how theory translates into practical preferences and legitimate concerns among the potential end-users (crew members).

4.6 Analytical Approach

Quantitative data was analyzed using descriptive statistics, including counts and percentages for each response option. Comparative analyses can be conducted by segmenting responses by experience level, AI familiarity, or scenario. For open-ended responses, automated sentiment and thematic analysis were applied, categorizing comments by sentiment (positive, neutral, negative) and extracting key themes such as trust development, accountability, safety, and human-AI interface design. The analytical approach allows for both high-level summary and detailed subgroup analysis.

4.7 Ethical Considerations

Participation was voluntary, with informed consent obtained at the outset. Respondents were informed of their right to withdraw at any time. The survey was anonymous, and no personally identifiable information was collected. Data is stored securely and used solely for research purposes. The study design adhered to ethical guidelines for research involving human subjects, ensuring respect for participant autonomy and data privacy.

4.8 Limitations

"Astronauts represent a unique population of healthy adults who have been carefully screened and selected based on their education, physical and mental health, and physical fitness" (Hupfeld et al., 2021).

The study's sample size (n=27) limits the generalizability of the findings; however, the total sample size of the target population was limited to the number of active astronauts currently employed at NASA and the number of analog astronauts currently employed at NASA. The voluntary and web-based distribution may introduce selection bias, favoring individuals who had the time in their training schedules to take the survey, as well as those who have a particular interest in AI. Self-reported data are subject to recall and social desirability biases, although these are less likely due to the anonymity. The absence of demographic variables restricts the analysis of potential confounding factors; however, the unavailability of this particular data was paramount in maintaining crew anonymity.

CHAPTER 5: EXPERIMENT AND RESULTS

5.1 Introduction

This chapter presents the findings from the experiment designed to gain insight into how astronauts and analog astronauts think artificial intelligence should be implemented in space systems for long duration spaceflight. These results will be combined with qualitative feedback from an industry expert to develop an overall interpretation of findings which will be described in Chapter 6. Appendix A provides the questions given to the astronauts and analog astronauts (participants) in a survey format. Appendix B provides the additional feedback provided by the participants for the open-ended question at the end of the survey.

5.2 Results

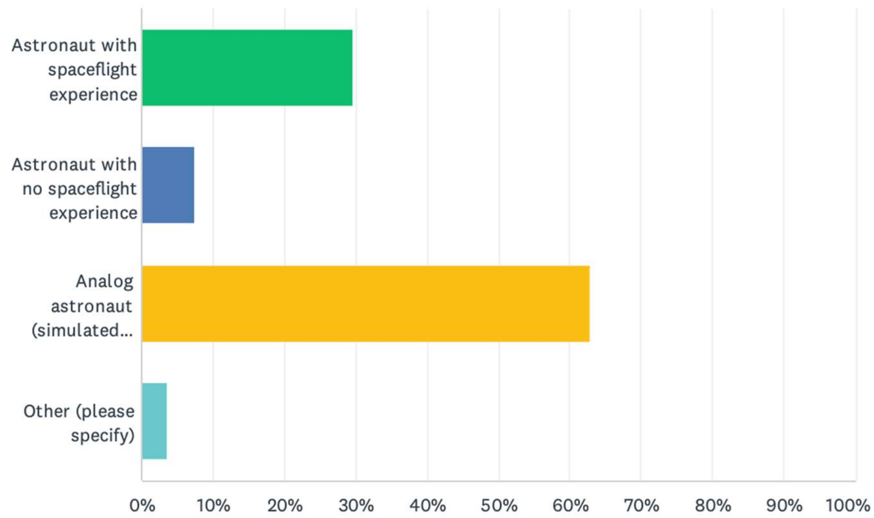
The survey, titled "Usability of Artificial Intelligence in Spacecraft for Long Duration Missions," collected responses from 27 participants, all of whom consented to participate. The data was gathered via a single web link collector. It is worth noting that 27 participants completed the first 5 questions, however only 25 participants completed the entire survey. For purposes of calculating statistical significance, 25 will be used as the total participant number.

5.2.1 Participant Experience Levels

The first section of the survey asked questions to expose the experience level of the participants. This includes their experience with spaceflight, using AI in general, and overall preferred role of AI in decision making. Results of the individual questions are as follows:

Q2 Experience Level:

Answered: 27 Skipped: 0

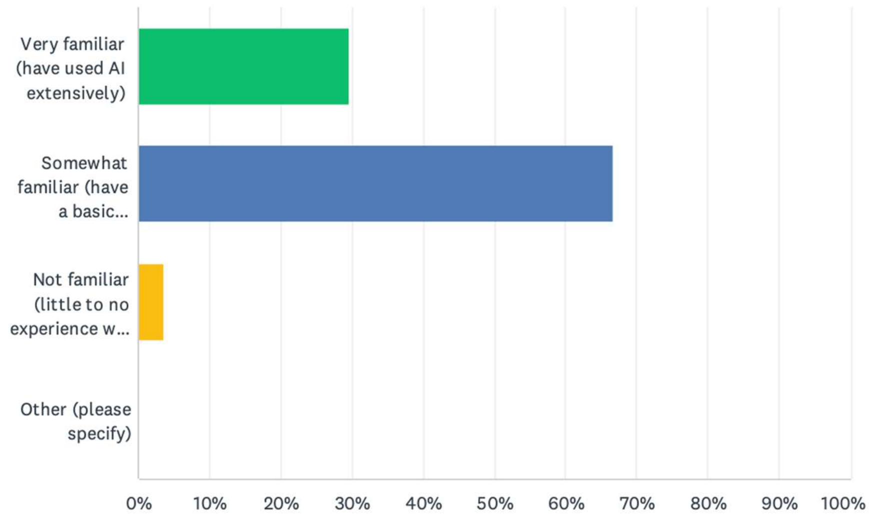


ANSWER CHOICES	RESPONSES	
Astronaut with spaceflight experience	29.63%	8
Astronaut with no spaceflight experience	7.41%	2
Analog astronaut (simulated mission experience)	62.96%	17
Other (please specify)	3.70%	1
Total Respondents: 27		

Figure 1. Results of Question #2

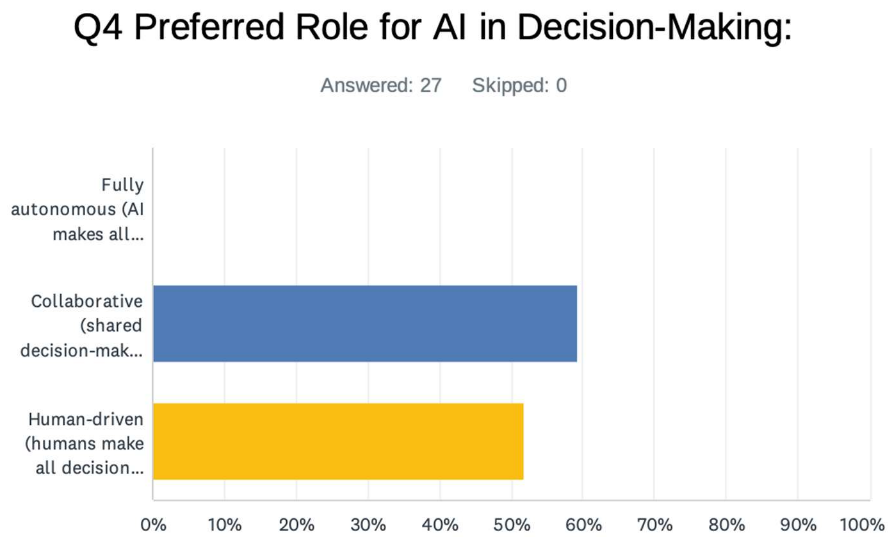
Q3 Familiarity with AI Systems:

Answered: 27 Skipped: 0



ANSWER CHOICES	RESPONSES	
Very familiar (have used AI extensively)	29.63%	8
Somewhat familiar (have a basic understanding of AI)	66.67%	18
Not familiar (little to no experience with AI)	3.70%	1
Other (please specify)	0.00%	0
Total Respondents: 27		

Figure 2. Results of Question #3

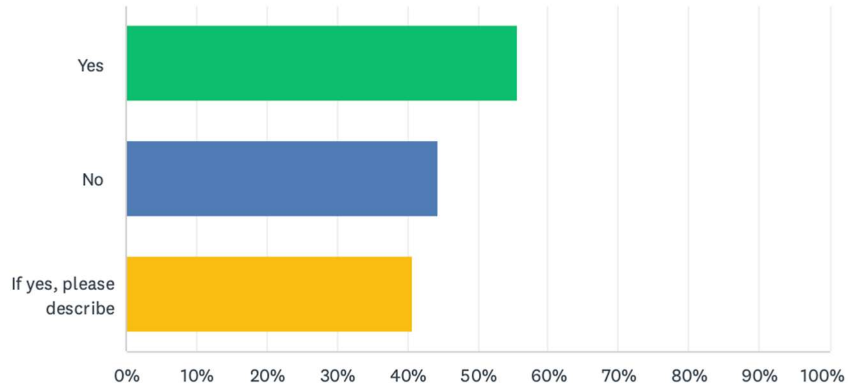


ANSWER CHOICES	RESPONSES	
Fully autonomous (AI makes all decisions)	0.00%	0
Collaborative (shared decision-making with AI)	59.26%	16
Human-driven (humans make all decisions with AI as a tool)	51.85%	14
Total Respondents: 27		

Figure 3. Results of Question #4

Q5 Previous Experience with AI or Automation in High-Stakes Settings:

Answered: 27 Skipped: 0



ANSWER CHOICES	RESPONSES	
Yes	55.56%	15
No	44.44%	12
If yes, please describe	40.74%	11
Total Respondents: 27		

Figure 4. Results of Question #5

5.2.2 Discrete Choice Tasks

The second section of the survey used the method of discrete choice tasks to measure preferences of artificial intelligence involvement by presenting hypothetical scenarios. The scenarios increased in severity of potential poor outcome.

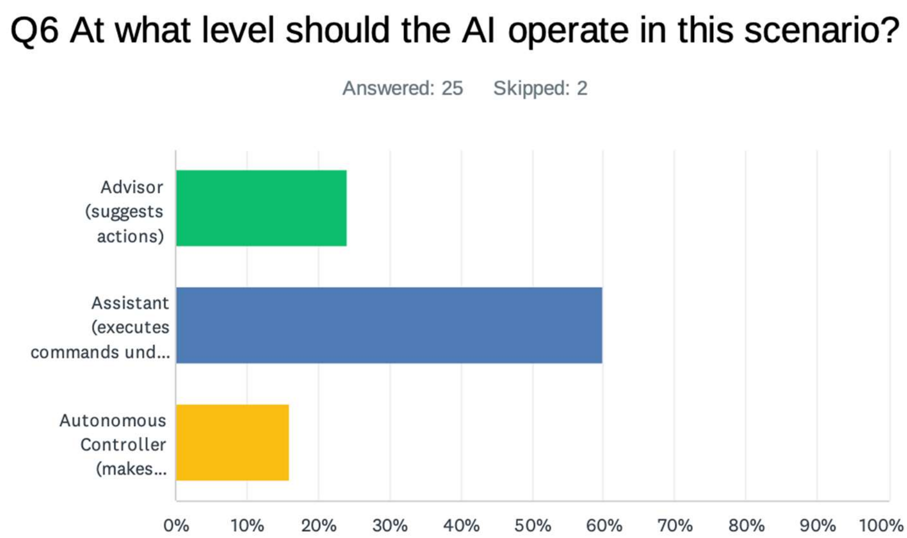
5.2.2.1 Scenario #1

The first scenario began with the following statement:

Scenario #1: Life Support

Maintaining Environmental Control and Life Support Systems (ECLSS). Setting optimal temperature and maintaining cabin pressure and humidity levels. Allocating rationing of food and water supplies to support mission duration and crew health.

The survey responses were as follows:

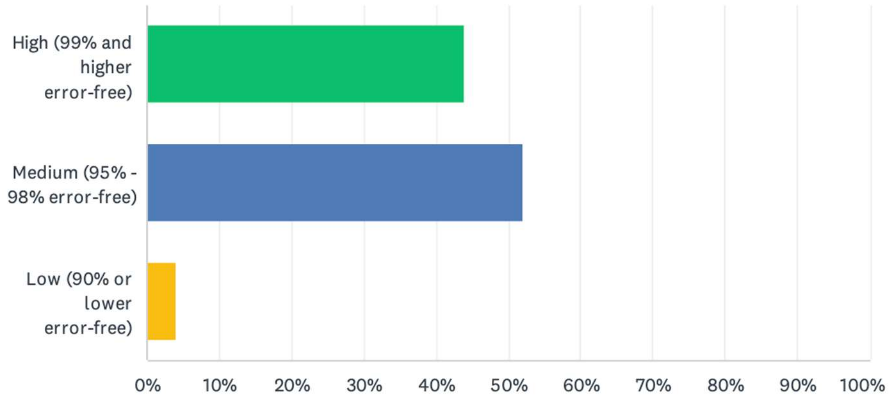


ANSWER CHOICES	RESPONSES	
Advisor (suggests actions)	24.00%	6
Assistant (executes commands under supervision)	60.00%	15
Autonomous Controller (makes independent decisions)	16.00%	4
TOTAL		25

Figure 5. Results of Question #6

Q7 What is the level of reliability of the AI needed for this scenario?

Answered: 25 Skipped: 2

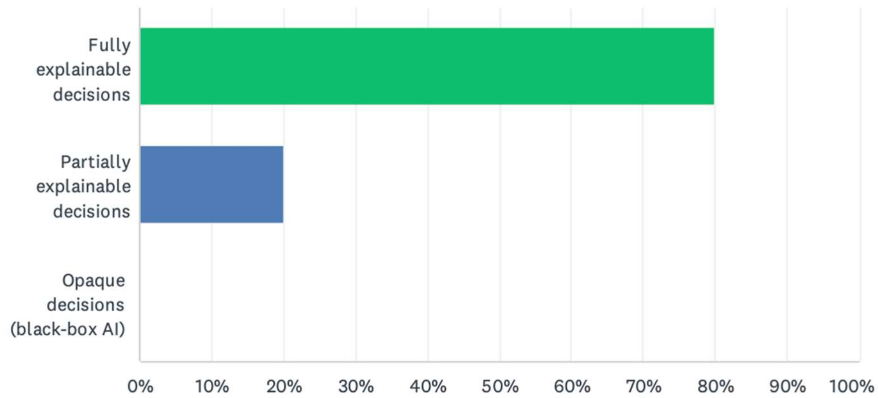


ANSWER CHOICES	RESPONSES	
High (99% and higher error-free)	44.00%	11
Medium (95% - 98% error-free)	52.00%	13
Low (90% or lower error-free)	4.00%	1
TOTAL		25

Figure 6. Results of Question #7

Q8 What level of transparency should the AI have in this scenario?

Answered: 25   Skipped: 2

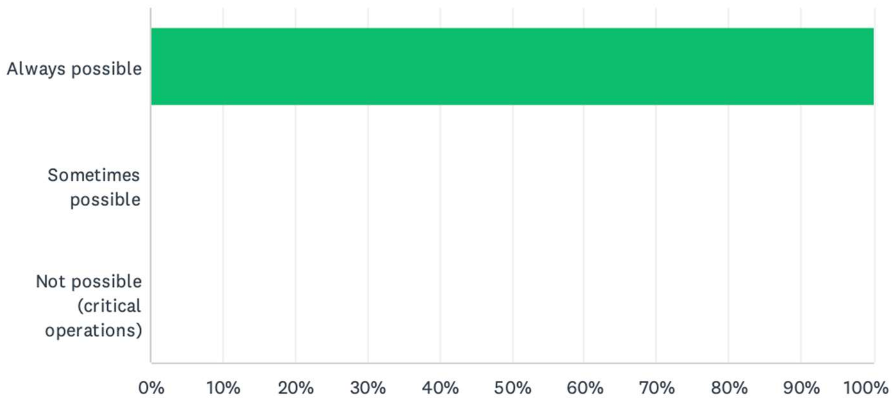


ANSWER CHOICES	RESPONSES	
Fully explainable decisions	80.00%	20
Partially explainable decisions	20.00%	5
Opaque decisions (black-box AI)	0.00%	0
TOTAL		25

Figure 7. Results of Question #8

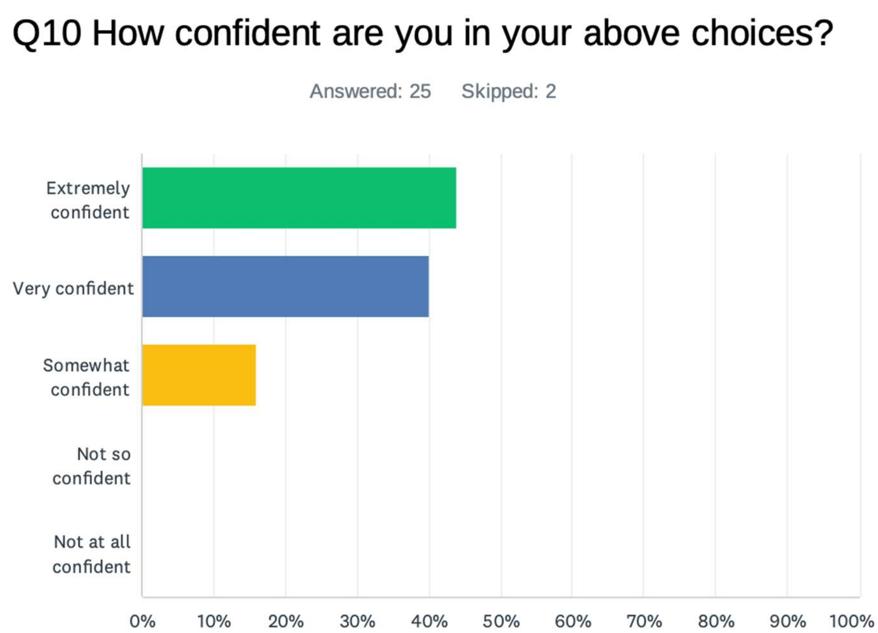
Q9 What level of human override capability should there be in this scenario?

Answered: 25 Skipped: 2



ANSWER CHOICES	RESPONSES	
Always possible	100.00%	25
Sometimes possible	0.00%	0
Not possible (critical operations)	0.00%	0
TOTAL		25

Figure 8. Results of Question #9



ANSWER CHOICES	RESPONSES	
Extremely confident	44.00%	11
Very confident	40.00%	10
Somewhat confident	16.00%	4
Not so confident	0.00%	0
Not at all confident	0.00%	0
TOTAL		25

Figure 9. Results of Question #10

5.2.2.2 Scenario #2

The second scenario began with the following statement:

Scenario #2: Routine Tasks

Performing routine tasks such as:

- Propellant level checks

- Trajectory accuracy checks

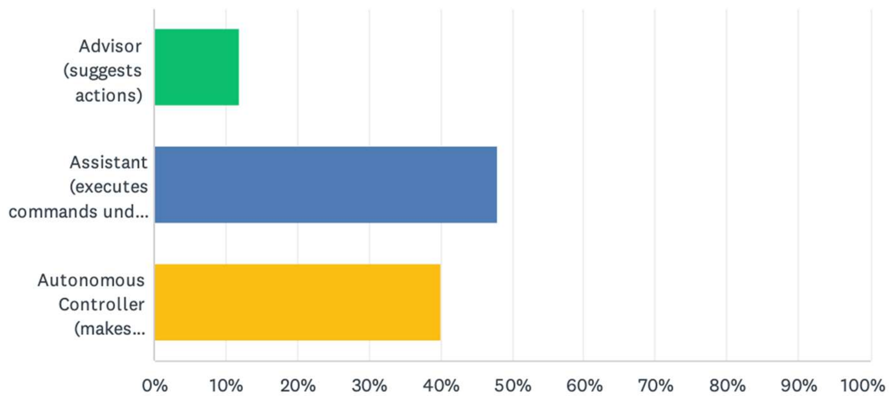
- Equipment functionality checks

- Visually inspecting spacecraft interior and exterior for potential issues

The survey responses were as follows:

Q11 At what level should the AI operate in this scenario?

Answered: 25 Skipped: 2

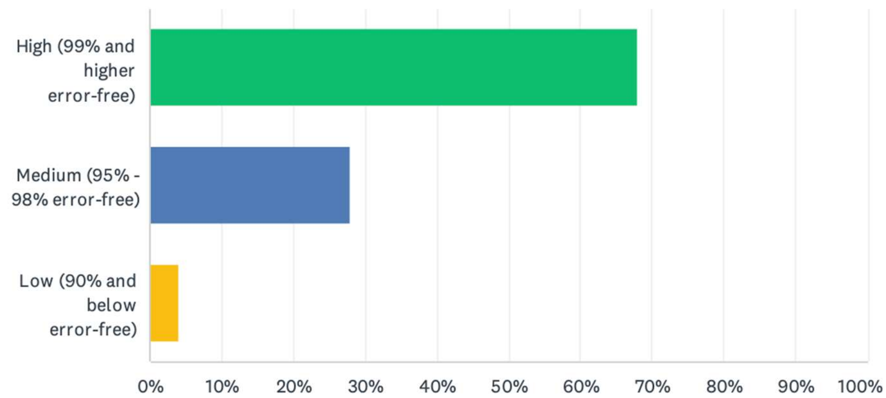


ANSWER CHOICES	RESPONSES	
Advisor (suggests actions)	12.00%	3
Assistant (executes commands under supervision)	48.00%	12
Autonomous Controller (makes independent decisions)	40.00%	10
TOTAL		25

Figure 10. Results of Question #11

Q12 What is the level of reliability of the AI needed for this scenario?

Answered: 25 Skipped: 2

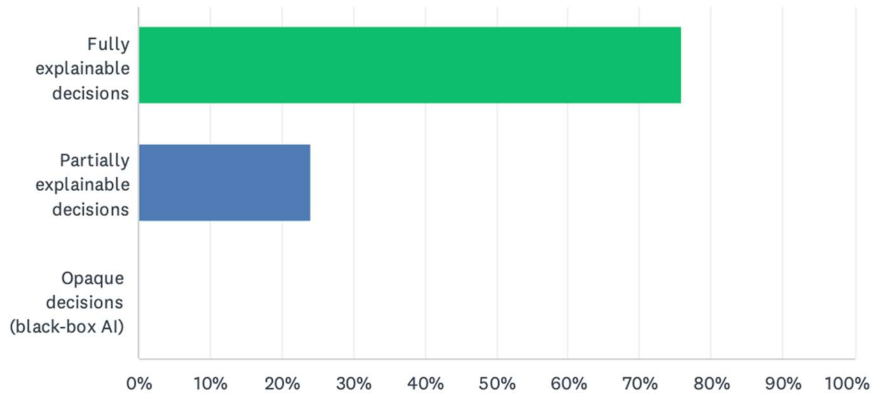


ANSWER CHOICES	RESPONSES	
High (99% and higher error-free)	68.00%	17
Medium (95% - 98% error-free)	28.00%	7
Low (90% and below error-free)	4.00%	1
TOTAL		25

Figure 11. Results of Question #12

Q13 What level of transparency should the AI have in this scenario?

Answered: 25 Skipped: 2

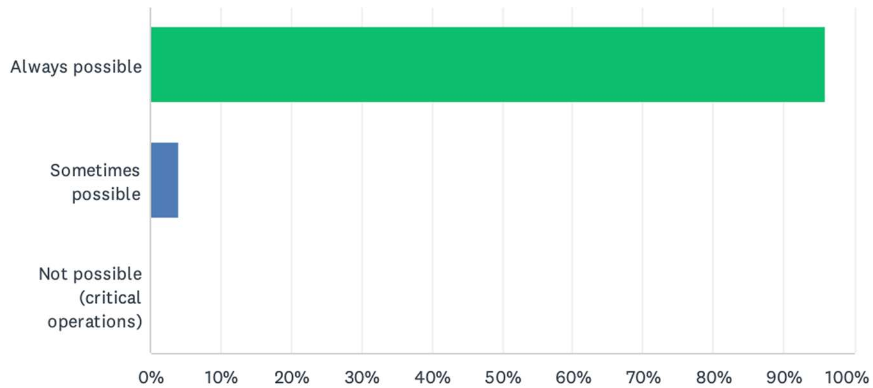


ANSWER CHOICES	RESPONSES	
Fully explainable decisions	76.00%	19
Partially explainable decisions	24.00%	6
Opaque decisions (black-box AI)	0.00%	0
TOTAL		25

Figure 12. Results of Question #13

Q14 What level of human override capability should there be in this scenario?

Answered: 25 Skipped: 2

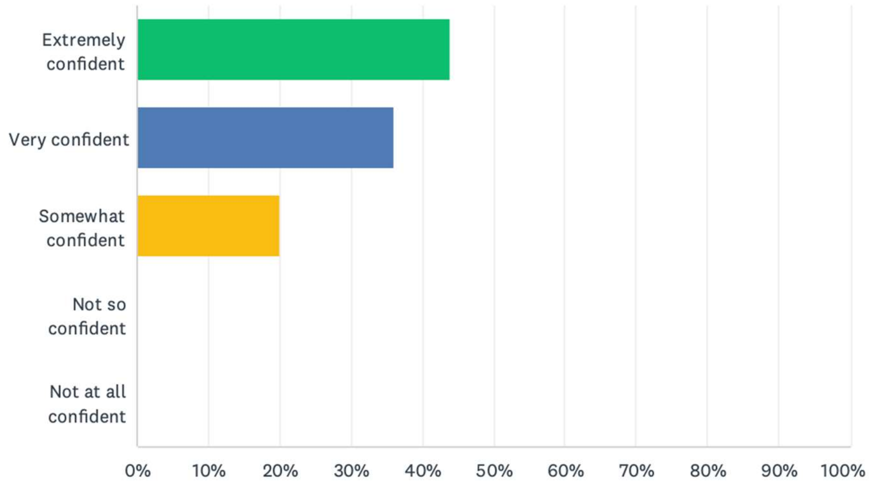


ANSWER CHOICES	RESPONSES	
Always possible	96.00%	24
Sometimes possible	4.00%	1
Not possible (critical operations)	0.00%	0
TOTAL		25

Figure 13. Results of Question #14

Q15 How confident are you in your above choices?

Answered: 25   Skipped: 2



ANSWER CHOICES	RESPONSES	
Extremely confident	44.00%	11
Very confident	36.00%	9
Somewhat confident	20.00%	5
Not so confident	0.00%	0
Not at all confident	0.00%	0
TOTAL		25

Figure 14. Results of Question #15

5.2.2.3 Scenario #3

The third scenario began with the following statement:

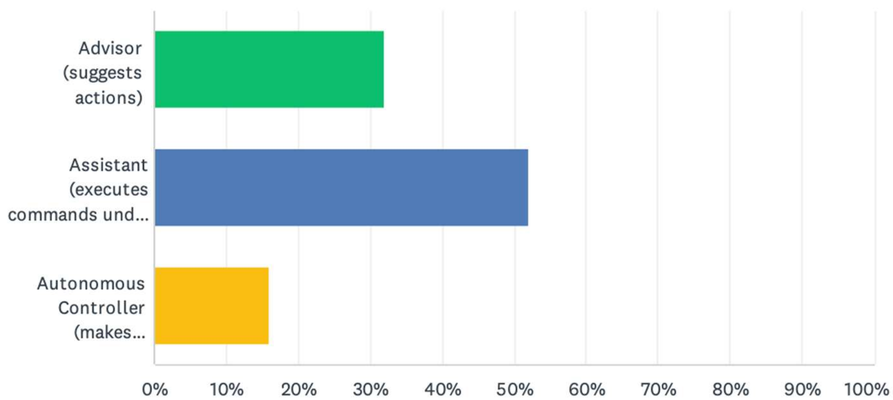
Scenario #3: Emergency (non-critical)

CO2 levels begin to rise in the spacecraft due to a malfunction with the primary CO2 scrubber. Backup scrubber is activated but the primary scrubber needs troubleshooting, evaluation, and potential fixes. AI could be used for the troubleshooting/evaluation/fix.

The survey responses were as follows:

Q16 At what level should the AI operate in this scenario?

Answered: 25 Skipped: 2

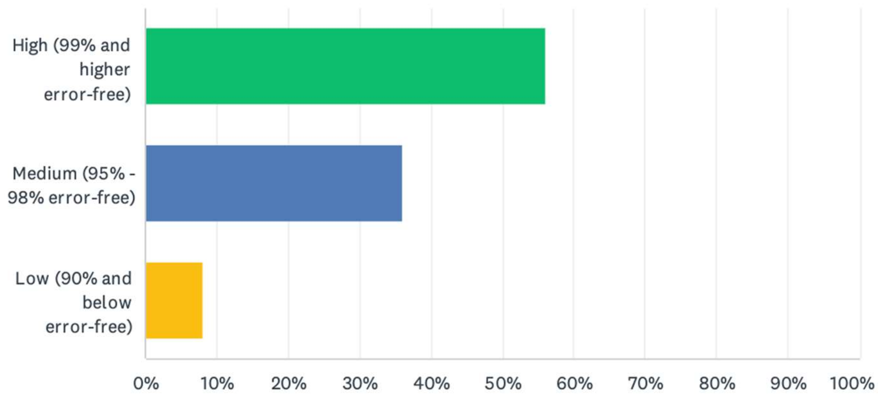


ANSWER CHOICES	RESPONSES	
Advisor (suggests actions)	32.00%	8
Assistant (executes commands under supervision)	52.00%	13
Autonomous Controller (makes independent decisions)	16.00%	4
TOTAL		25

Figure 15. Results of Question #16

Q17 What is the level of reliability of the AI needed for this scenario?

Answered: 25 Skipped: 2

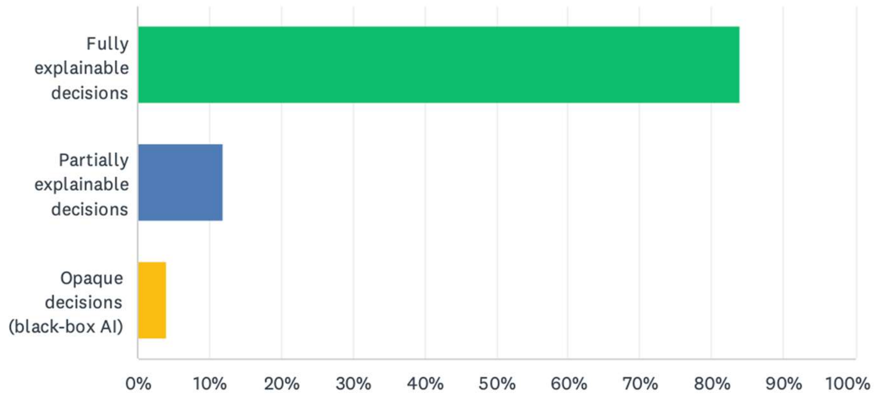


ANSWER CHOICES	RESPONSES	
High (99% and higher error-free)	56.00%	14
Medium (95% - 98% error-free)	36.00%	9
Low (90% and below error-free)	8.00%	2
TOTAL		25

Figure 16. Results of Question #17

Q18 What level of transparency should the AI have in this scenario?

Answered: 25 Skipped: 2

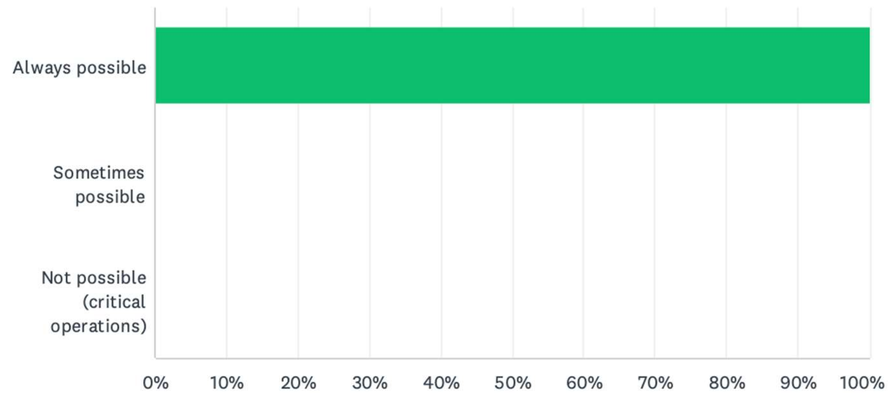


ANSWER CHOICES	RESPONSES	
Fully explainable decisions	84.00%	21
Partially explainable decisions	12.00%	3
Opaque decisions (black-box AI)	4.00%	1
TOTAL		25

Figure 17. Results of Question #18

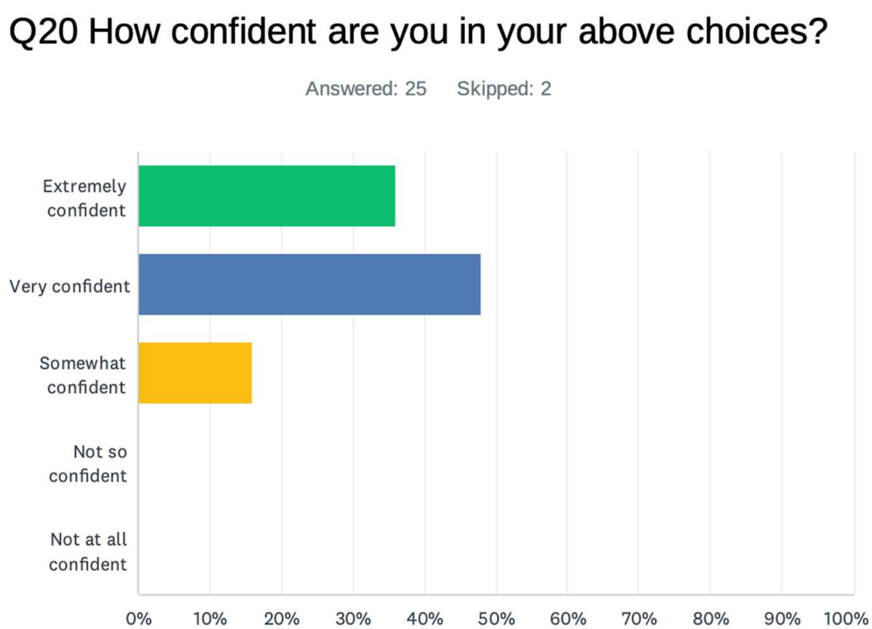
Q19 What level of human override capability should there be in this scenario?

Answered: 25 Skipped: 2



ANSWER CHOICES	RESPONSES	
Always possible	100.00%	25
Sometimes possible	0.00%	0
Not possible (critical operations)	0.00%	0
TOTAL		25

Figure 18. Results of Question #19



ANSWER CHOICES	RESPONSES	
Extremely confident	36.00%	9
Very confident	48.00%	12
Somewhat confident	16.00%	4
Not so confident	0.00%	0
Not at all confident	0.00%	0
TOTAL		25

Figure 19. Results of Question #20

5.2.2.4 Scenario #4

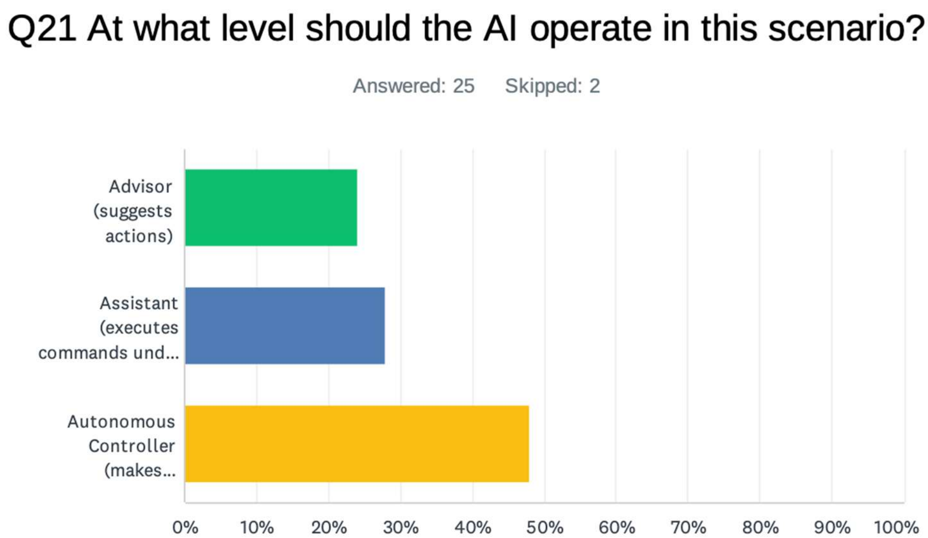
The fourth and final scenario began with the following statement:

Scenario #4: Emergency (Mission Critical)

During spaceflight, alarms sound indicating a sudden drop in cabin pressure. A micrometeorite has struck the spacecraft, puncturing a small but dangerous hole in the hull. Air is rapidly escaping and oxygen levels are dropping. If not addressed

immediately, crew risks losing breathable air, leading to asphyxiation or structural failure of the spacecraft.

The survey responses are as follows:

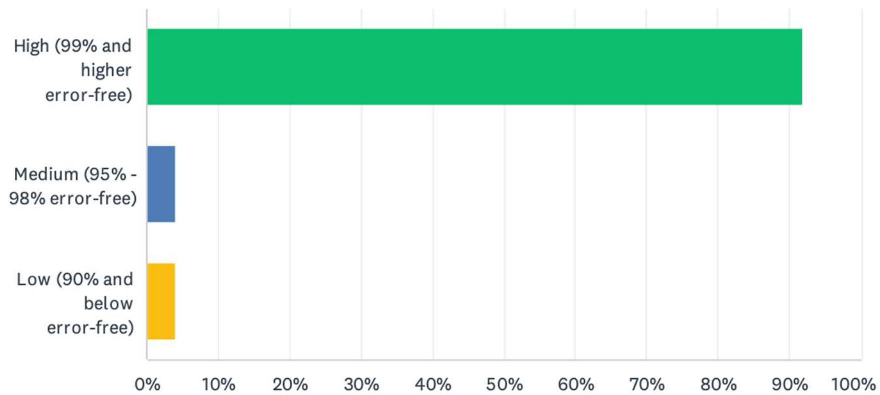


ANSWER CHOICES	RESPONSES	
Advisor (suggests actions)	24.00%	6
Assistant (executes commands under supervision)	28.00%	7
Autonomous Controller (makes independent decisions)	48.00%	12
TOTAL		25

Figure 20. Results of Question #21

Q22 What is the level of reliability of the AI needed for this scenario?

Answered: 25   Skipped: 2

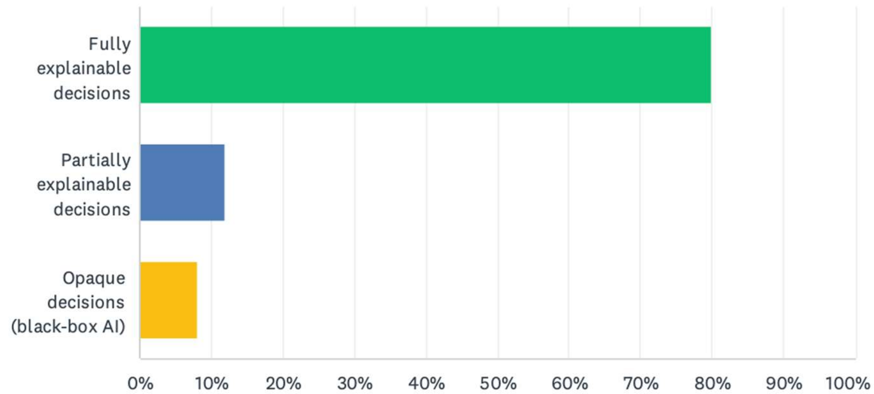


ANSWER CHOICES	RESPONSES	
High (99% and higher error-free)	92.00%	23
Medium (95% - 98% error-free)	4.00%	1
Low (90% and below error-free)	4.00%	1
TOTAL		25

Figure 21. Results of Question #22

Q23 What level of transparency should the AI have in this scenario?

Answered: 25 Skipped: 2

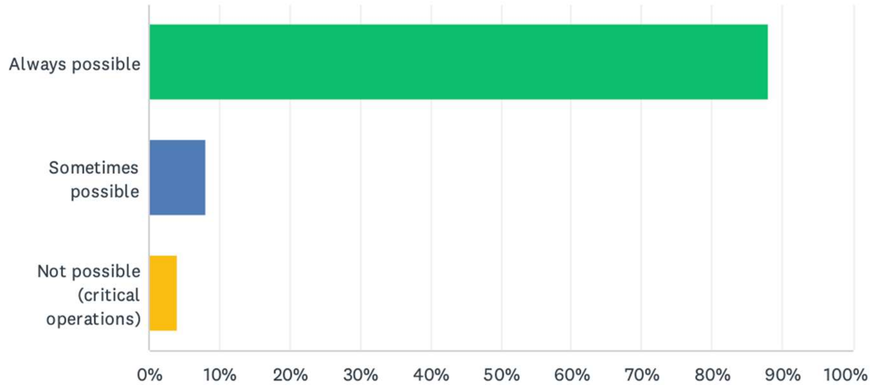


ANSWER CHOICES	RESPONSES	
Fully explainable decisions	80.00%	20
Partially explainable decisions	12.00%	3
Opaque decisions (black-box AI)	8.00%	2
TOTAL		25

Figure 22. Results of Question #23

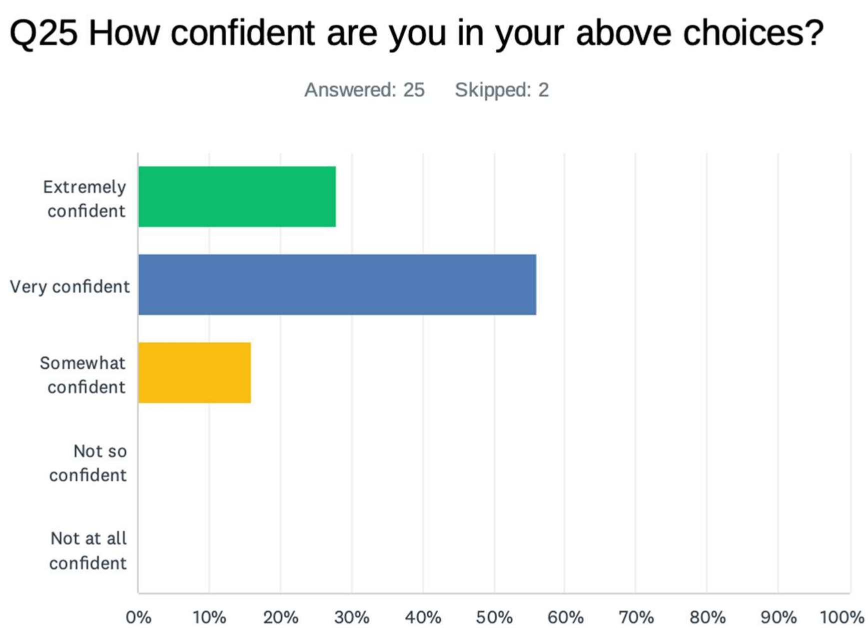
Q24 What level of human override capability should there be in this scenario?

Answered: 25 Skipped: 2



ANSWER CHOICES	RESPONSES	
Always possible	88.00%	22
Sometimes possible	8.00%	2
Not possible (critical operations)	4.00%	1
TOTAL		25

Figure 23. Results of Question #24



ANSWER CHOICES	RESPONSES	
Extremely confident	28.00%	7
Very confident	56.00%	14
Somewhat confident	16.00%	4
Not so confident	0.00%	0
Not at all confident	0.00%	0
TOTAL		25

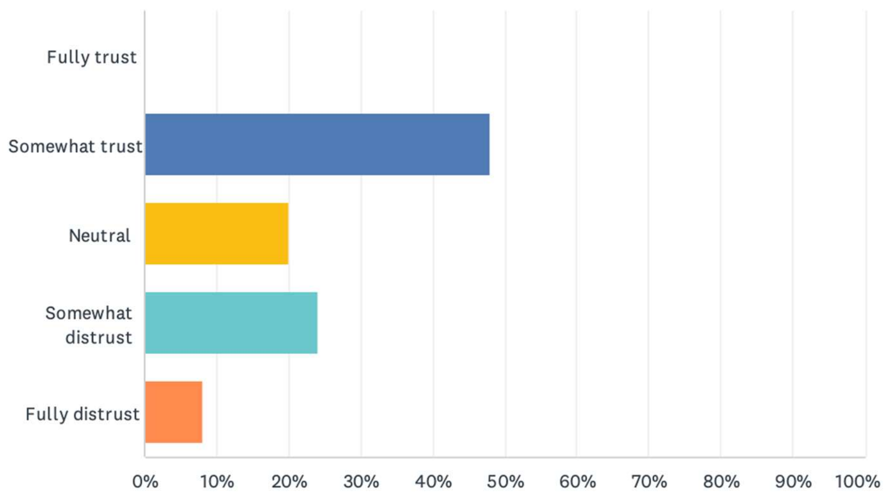
Figure 24. Results of Question #25

5.2.3 Additional Questions

The third section of the survey held additional generalized artificial intelligence questions for the participants. The results are as follows:

Q26 How much do you trust AI systems in high-stakes environments?

Answered: 25 Skipped: 2

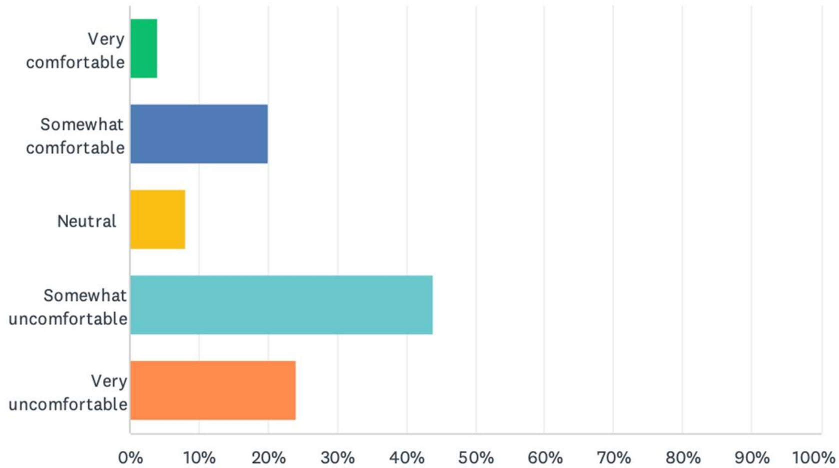


ANSWER CHOICES	RESPONSES	
Fully trust	0.00%	0
Somewhat trust	48.00%	12
Neutral	20.00%	5
Somewhat distrust	24.00%	6
Fully distrust	8.00%	2
TOTAL		25

Figure 25. Results for Question #26

Q27 How comfortable are you with AI making critical decisions without human involvement?

Answered: 25 Skipped: 2

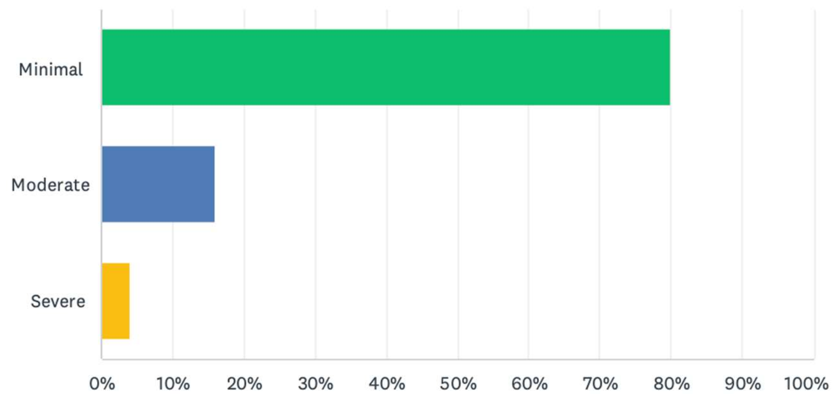


ANSWER CHOICES	RESPONSES	
Very comfortable	4.00%	1
Somewhat comfortable	20.00%	5
Neutral	8.00%	2
Somewhat uncomfortable	44.00%	11
Very uncomfortable	24.00%	6
TOTAL		25

Figure 26. Results for Question #27

Q28 What level of risk would you accept for an AI failure in life support systems?

Answered: 25 Skipped: 2

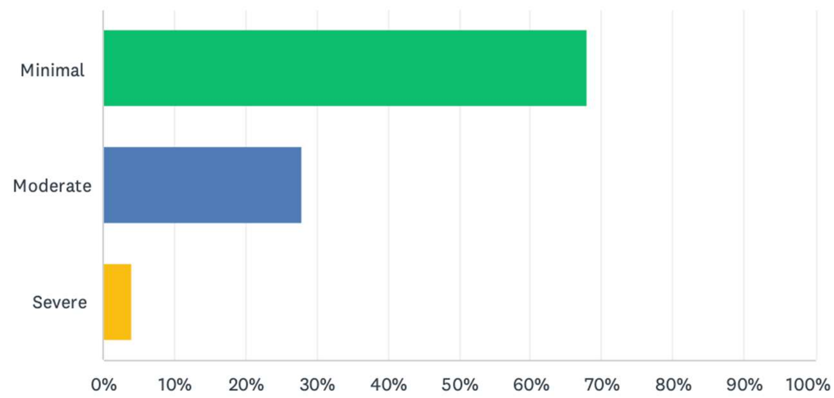


ANSWER CHOICES	RESPONSES	
Minimal	80.00%	20
Moderate	16.00%	4
Severe	4.00%	1
TOTAL		25

Figure 27. Results for Question #28

Q29 What level of risk would you accept for an AI failure in navigation and docking?

Answered: 25   Skipped: 2

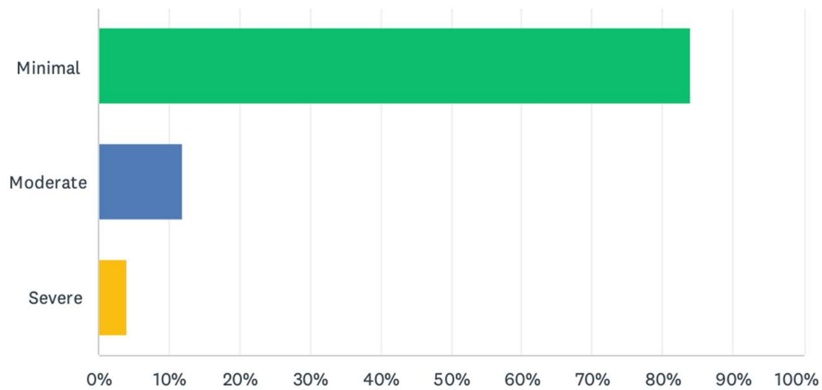


ANSWER CHOICES	RESPONSES	
Minimal	68.00%	17
Moderate	28.00%	7
Severe	4.00%	1
TOTAL		25

Figure 28. Results for Question #29

Q30 What level of risk would you accept for an AI failure in an emergency response situation?

Answered: 25 Skipped: 2

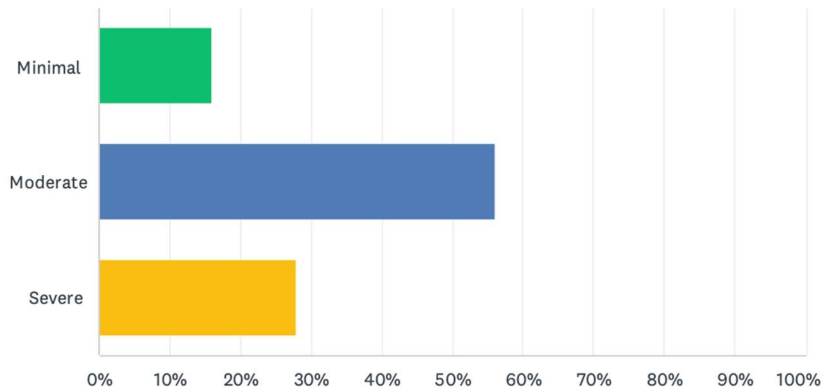


ANSWER CHOICES	RESPONSES	
Minimal	84.00%	21
Moderate	12.00%	3
Severe	4.00%	1
TOTAL		25

Figure 29. Results for Question #30

Q31 What level of risk would you accept for an AI failure in a scientific experiment?

Answered: 25 Skipped: 2

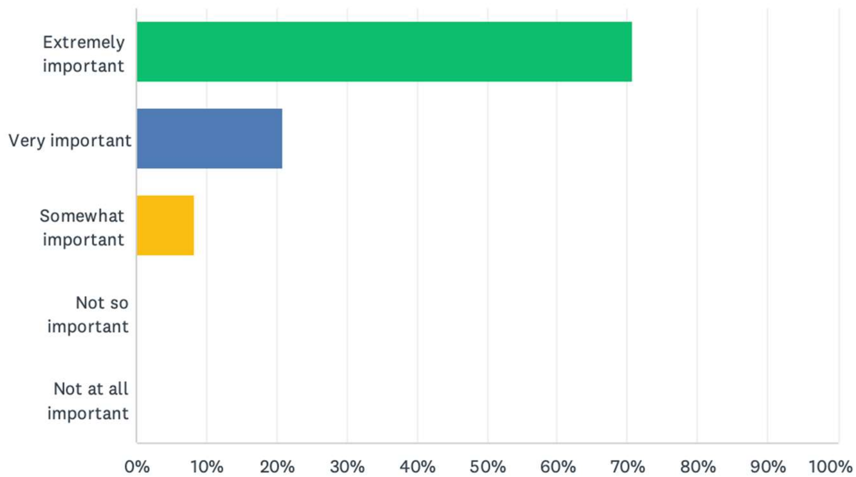


ANSWER CHOICES	RESPONSES	
Minimal	16.00%	4
Moderate	56.00%	14
Severe	28.00%	7
TOTAL		25

Figure 30. Results for Question #31

Q32 How important is it for an AI system to explain its decisions?

Answered: 24 Skipped: 3

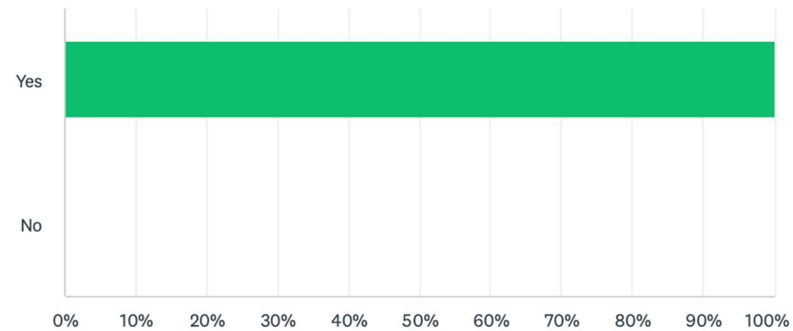


ANSWER CHOICES	RESPONSES	
Extremely important	70.83%	17
Very important	20.83%	5
Somewhat important	8.33%	2
Not so important	0.00%	0
Not at all important	0.00%	0
TOTAL		24

Figure 31. Results for Question #32

Q33 Would you want to take part in the development and training of the AI system being used on a spacecraft you will fly on?

Answered: 25 Skipped: 2



ANSWER CHOICES	RESPONSES	
Yes	100.00%	25
No	0.00%	0
TOTAL		25

Figure 32. Results of Question #33

5.2.4 Final Thoughts

The final section of the survey allowed the participant an open-ended response to add any final thoughts they would like to share regarding AI integration in spacecraft for long duration missions. The responses to this question are included in Appendix B.

5.3 Interesting and Notable Trends

After analyzing the survey results from a top-down approach, starting with a high-level view and drilling down into specific details, the following notable trends were identified:

1. The majority of participants have experience as analog astronauts (62.96%), and most are at least somewhat familiar with AI systems (66.67%).
2. When it comes to the preferred role for AI in decision-making, most favor a collaborative approach (59.26%) or human-driven with AI as a tool (51.85%), while no one prefers fully autonomous AI.
3. Across multiple scenarios, the most popular AI operation level is as an "Assistant" (executes commands under supervision), but there is some variation depending on the scenario.
4. High reliability (99%+ error-free) is consistently valued, especially in later scenarios as the severity of the simulated scenarios increased (up to 92% in one case).
5. Transparency is highly valued, as the vast majority want AI decisions to be fully explainable (up to 84%).
6. Human override is seen as essential, with nearly all respondents wanting it to always be possible.
7. Trust in AI is moderate: most "somewhat trust" AI in high-stakes environments, but few "fully trust" it.
8. Most are uncomfortable with AI making critical decisions without human involvement.
9. Minimal risk is the most acceptable for AI failures in life support, navigation, and emergency response, but more are willing to accept moderate or severe risk in scientific experiments.

10. Nearly all respondents (100%) would want to participate in the development and training of the AI system they would use.

5.4 Insights from Qualitative Feedback

Qualitative feedback was collected through an open-ended question asking participants to share additional comments regarding AI integration in spacecraft for long duration missions. The responses to this question were analyzed for both sentiment and themes.

5.4.1 Sentiment Analysis

The majority of comments were seen to be neutral (70%), with a smaller proportion expressing negative (20%) and positive (10%) sentiments. This suggests that while most participants provided balanced or factual feedback, there is a notable minority with concerns or optimism about AI integration.

5.4.2 Thematic Analysis

The open-ended feedback consisted of a wide range of themes. The key themes that have been identified are:

- AI trust development
- AI accountability mechanisms
- Emergency response support
- AI assistance in operations
- AI safety constraints
- AI training processes
- Risk mitigation strategies
- AI decision explainability
- AI workflow integration
- Experience-based evaluation

- Performance evaluation metrics
- System verification procedures
- Human-AI interface design
- Human oversight functions
- Engineering knowledge retention

These themes indicate that participants are particularly interested in how AI can be made trustworthy, accountable, and safe, as well as how it can support operations and emergency responses. There is also a focus on the importance of explainability, human oversight, and robust training and verification processes. Each of these themes represent an area of study that should be of interest to the research community to further prepare for AI integration into spacecraft.

5.5 Limitations and Challenges

5.5.1 Statistical Significance

For the purposes of generalization, it is important to note the number of people in the target population compared to the number of people who participated in the study. The target population is the NASA analog astronaut office and the NASA astronaut office. The survey was sent to 50 people within the analog astronaut office and 73 people within the NASA astronaut office. To calculate the statistical significance, the following equation is used:

$$E = Z * \sqrt{\frac{p(1-p)}{n}} * \sqrt{\frac{N-n}{N-1}}$$

Figure 33. Margin of Error Equation

The target population for this study consisted of a very finite and specialized group of people (N=123). A total of 25 participants completed the survey, representing approximately 20% of the target population. Given the small and bounded population size, a finite population correction was applied when estimating precision. Assuming a conservative population proportion of 0.5, the achieved margin of error at the 95% confidence level was approximately ±17.6%. While this margin of error exceeds that typically associated with large-population surveys, it is consistent with expectations for research involving small, highly specialized populations where participant access is inherently constrained, especially with the known rigor of astronaut training schedules. The resulting sample size is sufficient to support exploratory analysis, identify generalized trends, and inform requirement development and hypothesis generation.

5.5.2 Voluntary Responses

Participation in this study was entirely voluntary, consistent with ethical research practices for human-subjects research involving NASA personnel. As a result, the sample reflects a self-selected subset of the target population, which may introduce response bias if individuals with stronger opinions or greater interest in the topic were more likely to participate. However, voluntary participation was necessary due to the specialized nature of the population and the operational demands placed on potential participants (crew schedules, etc.). This approach is typical for research involving a constrained, high-expertise population and is appropriate for exploratory objectives aimed at capturing individual perspectives, informing system requirements, and identifying human-factors considerations rather than producing statistically representative estimates of the broader population.

CHAPTER 6: ANALYSIS AND DISCUSSION

6.1 Introduction

This chapter will analyze and discuss the survey results and expand on the key notable findings noted in section 5.3.

6.2 Interpretation of Key Findings

Respondents represent a range of experience levels, with the majority being analog astronauts (62.96%), followed by astronauts with spaceflight experience (29.63%). Most participants are somewhat familiar with AI systems (66.67%), and a significant portion have previous experience with AI or automation in high-stakes settings (55.56%). This range of participant background works to validate the responses as being capable of generalization among the target population of end-users for AI technology implementation in spacecraft.

Across various scenarios, respondents consistently favored AI operating as an assistant or advisor rather than as a fully autonomous controller. High reliability (99%+ error-free) and fully explainable decisions were strongly preferred, and there was overwhelming support for always allowing human override capability.

When it comes to trust and comfort, most respondents expressed only moderate trust in AI for high-stakes environments and were somewhat or very uncomfortable with AI making critical decisions without human involvement. Minimal risk was the most acceptable level for AI failures in life support, navigation, and emergency response, while moderate risk was more accepted for scientific experiments.

Thematic and sentiment analysis of open-ended comments revealed a predominance of neutral sentiment, with key themes including trust development, training processes, system verification, and risk mitigation.

6.2.1 Summary

The data highlights a cautious but open attitude toward AI integration in critical space missions. Respondents value transparency, reliability, and human oversight, indicating that successful AI adoption in this context will require robust safety, explainability, and collaboration mechanisms.

6.3 Comparison with Existing Literature

The respondent mix (high analog experience, moderate AI familiarity) is consistent with the broader deep-space framing in NASA human exploration work, where operations are increasingly shaped by autonomy and constrained communication, but humans remain central to mission outcomes. NASA's exploration roadmap and HRP framing emphasize that deep-space introduces unique constraints (latency, isolation, high workload) that make autonomy attractive but also heighten the need for human-centered approaches.

The key finding that no one prefers fully autonomous AI matches a recurring theme across human-AI systems and space autonomy discussions, which is autonomy is valuable, but authority allocation and human responsibility remain largely valuable, especially under uncertainty and safety-critical conditions. NASA's autonomy work in operations is positioned as enabling and augmenting mission execution rather than removing humans from accountability (NASA, 2021). In human-centered design terms, this aligns with the need for user control and freedom (Nielsen, 2024) and with trustworthiness discussions that emphasize bounded behavior and assured oversight for AI used in high-consequence settings (Rößler et al., 2022).

The literature strongly supports a pattern that has been identified from the survey results, which shows as consequence increases, acceptable error rates shrink dramatically. Trustworthiness and AI testing guidance explicitly treat high-consequence contexts as requiring robust verification, validation, and assurance practices that go beyond average performance metrics (Rößler et al., 2022). In space operations specifically, resilient systems engineering emphasizes bounded failure behavior and credible recovery pathways (Day et al., 2015). The survey adds an operator-centered quantification where respondents expect more reliability as severity increases, reinforcing the need for graded assurance (where the level of testing rigor is applied in a scaling manner according to the criticality of the operational situation).

When it comes to the results on transparency (up to 84% wanting fully explainable AI) the results align with both usability and trustworthiness literature Nielsen's heuristics (visibility of system status, match between system and real world, help users recognize/diagnose/recover from errors) predict that AI that cannot explain its actions will be rejected when users must supervise, diagnose, or intervene (Nielsen, 2024). Trustworthy AI requires traceability, testability, and evidence that outputs are justifiable in context, especially when stakes are high (Rößler et al., 2022). The survey data strengthens this by showing explainability isn't simply a "nice to have", it's central to the users' acceptance of AI decision-making in the spaceflight environment.

6.4 Implications

Risk is the key feature that must be mitigated when implementing AI systems in spacecraft. Mitigating this unique risk must involve rigorous verification and validation practices, AI functional fitness assessments, and a segmented approach for AI application.

6.4.1 Verification and Validation of AI

The strong emphasis on near-perfect reliability, transparency, and always-available human override indicates that traditional software verification and validation (V&V) approaches may be insufficient for AI systems intended for deep-space operations. Respondents' reluctance to accept autonomous decision-making, particularly in high-stakes scenarios, suggests that AI V&V must extend beyond performance metrics to include behavioral predictability, explainability, and failure mode characterization. Verification activities should therefore explicitly assess whether AI systems behave within expected bounds under off-nominal and degraded conditions, and whether their output remains explainable to human operators. Validation must include human-in-the-loop testing, ensuring that operators can understand, supervise, and override AI actions effectively, even under time pressure and large cognitive load. These findings imply that confidence in AI for spaceflight will be driven by demonstrated controllability (human override capabilities) and interpretability (AI explaining its decisions).

6.4.2 AI Functional Fitness Assessment

The consistent preference for AI operating in an “Assistant” role, combined with moderate trust levels and discomfort with unsupervised critical decisions, suggests the need for an AI Functional Fitness Assessment which should be conducted in parallel to crew or system readiness evaluations. Such an assessment would evaluate not only technical performance, but also whether an AI system is fit for a specific operational context, mission phase, and level of human involvement. Functional fitness criteria may include reliability thresholds appropriate to task criticality, the quality and timeliness of explanations provided, the effectiveness of human override mechanisms, and the system’s ability to adapt to changing mission conditions without eroding operator trust. Functional fitness should be scenario-dependent rather than absolute (such as how the survey has been designed), to observe variation in acceptable AI roles across different operational contexts.

6.4.3 Segmentation of Spaceflight Landscape for AI Application

The differentiated risk tolerance observed across system functions indicates that AI integration in spaceflight should be segmented by domain, criticality, and consequence of failure, rather than implemented uniformly across the vehicle or mission. Minimal risk tolerance for life support, navigation, and emergency response implies that AI in these domains should remain tightly supervised, with conservative autonomy levels and explainability requirements. In contrast, the higher willingness to accept moderate or severe risk in scientific experimentation suggests that these domains may serve as early adoption areas for more advanced or adaptive AI capabilities (perhaps as a testing ground, so to speak). This segmentation approach allows AI maturity to progress across the spaceflight landscape, aligning autonomy levels with both technical risk and human risk level acceptance. Such a strategy supports incremental trust building, targeted V&V practices, and clearer prioritization of where highly trustworthy AI is essential versus where exploratory use is acceptable.

6.5 Recommendations for Future Research

6.5.1 Hallucinations

One critical area of research that is not explored in this experiment is the concept of AI hallucinations. A key limitation of this study is that it captures operator perceptions and preferences regarding AI use, rather than directly measuring AI failure behaviors such as hallucinations under operational conditions. While respondents consistently emphasized reliability, transparency, and human override, the mechanisms by which AI hallucinations emerge and affect decision-making in spaceflight contexts remain insufficiently characterized. Future work should examine hallucination frequency, detectability, and operational impact across mission-relevant scenarios, particularly under conditions of delayed communication, incomplete data, and elevated cognitive load on the crew members. Research should prioritize human-in-the-loop evaluations to assess whether operators can recognize hallucinations in real time, interpret AI confidence and explanations appropriately, and intervene effectively before safety margins are compromised. Addressing these gaps is critical to developing trust frameworks that treat hallucinations as a distinct, safety-relevant failure mode rather than as generalized model error.

6.5.2 Human-AI Interface Design

Feedback from the open-ended survey question strongly suggests that acceptance of AI depends as much on the human interface and workflow integration as on raw performance. Future studies should evaluate different interface designs, explanation presentations, and interaction protocols to understand how they influence situational awareness, workload, and trust. Research should also explore how AI can integrate seamlessly into existing crew routines rather than introducing parallel or disruptive workflows.

6.5.3 Explainability, Logging, and Post-Hoc Accountability

Per the survey results, while explainability was widely valued, participants noted that real-time explainability may not always be feasible. Future research should investigate graded explanation strategies, including minimal real-time explanations paired with more detailed post-event logs. Work is needed to determine what forms of logging best support trust, forensic analysis, training, and accountability without overloading crews during operations.

6.5.4 Open-Ended Interview Questions

Due to the rigid nature of classic survey questions, there is a desire to go beyond the initial questions and dive deeper into the intentions behind initial responses. Another survey could be designed to have additional open-ended questions immediately following the quantitative questions. Including open-ended questions following quantitative survey questions is beneficial because it provides essential contextual insight into why respondents selected particular ratings or preferences, especially in complex, high-stakes domains such as human–AI interaction. While quantitative responses capture trends and distributions, open-ended feedback helps uncover underlying assumptions, mental models, and conditional reasoning that may not be visible in scaled responses alone. This qualitative input can reveal unanticipated concerns, clarify ambiguities in interpretation, and support the development of more actionable requirements by linking basic outcomes to operator experience and intent. In exploratory studies with small, specialized populations, open-ended responses also strengthen construct validity by ensuring that quantitative findings are grounded in participant meaning rather than researcher inference.

For example, in the survey given for this research there are quantitative questions that could have been followed by open-ended questions, such as:

Table 2. Examples of Follow-on Open-ended Survey Questions

Quantitative Survey Question	**Follow-on Open-ended Questions**
Preferred Role for AI in Decision-Making	Why do you prefer this role? Is there anything from a design or V&V standpoint that could be implemented to alter your choice?

At what level should the AI operate in this scenario?	Why do you prefer this level? What would be required for you to trust the AI to operate at a higher trust level?
How important is it for an AI system to explain its decisions?	What drives the need for explainability?
How comfortable are you with AI making critical decisions without human involvement?	What from a system design standpoint would make you more comfortable with AI making critical decisions without human involvement?
What level of risk would you accept for an AI failure in life support systems?	What additional training or V&V could be applied to the systems engineering process to alleviate the fear of AI failure?
Would you want to take part in the development and training of the AI system being used on a spacecraft you will fly on?	At what level would you want to be involved in the development and training? At what stage in the systems engineering process would you want your involvement to begin?
How confident are you in your above choices?	What is driving your confidence level? What, from a technical standpoint, could increase your confidence level?

Open-ended data also strengthen construct validity by anchoring survey metrics in participant language and lived experience. They provide traceability between measured outcomes and real-world implications, enabling more defensible interpretation of results. Qualitative responses humanize the data, giving voice to expert judgment and enabling richer synthesis with existing literature. When paired thoughtfully with quantitative questions, open-ended responses transform a survey from a measurement instrument into a discovery tool, particularly well-suited for early-stage research and complex sociotechnical systems. This is especially useful when researching AI, as there is such a vast array of thoughts on the development, trust, and application of such systems.

CHAPTER 7: CONCLUSION

AI is proving to be an integral part of managing the complexities of long-duration space missions. AI systems offer solutions for autonomous decision-making, resource optimization, and crew well-being, which are critical for mission success in environments where communication with Earth is significantly delayed and cognitive load is high. However, successful integration of AI requires addressing challenges related to system reliability, human interaction, and the critical balance between autonomy and human oversight. Future work should focus on advancing simulation techniques and developing AI systems that can operate effectively in space's unpredictable and harsh environment, as well as optimizing the trust between human and machine. Future work should also include research into AI hallucinations, and how that might impact the operational capabilities of AI, as well as affect the trust levels of the crew members who are using the AI. AI requires verification and validation, sensitivity analysis, integrated test and measurement, and automatic engagement with a human in the loop when the AI confidence (either/or self-reported by the AI or by the V/V processes, etc.) is low.

In the end, keeping the human in the loop is an integral part of the systems design process, and should allow for crew members to participate in the development and training of the AI. Just like how crew members assigned to the same flight/mission train together, so should the AI train together with them and be treated as another crew member.

References

Swenson, L. S., Grimwood, J. M., & Alexander, C. C. (1998). This new ocean: a history of Project Mercury. National Aeronautics and Space Administration, NASA History Office, Office of Policy and Plans.

National Aeronautics and Space Administration. (2021). *Autonomous Systems and Operations*. NASA. https://techport.nasa.gov/projects/32946.

Lopatovska, Irene et al. "Talk to Me: Exploring User Interactions with the AmazonAlexa." *Journal of librarianship and information science* 51.4 (2019): 984–997. Web.

Fanti, Lucrezia, Dario Guarascio, and Massimo Moggi. "From Heron of Alexandria to Amazon's Alexa: A Stylized History of AI and Its Impact on Business Models, Organization and Work." Economia e politica industriale 49.3 (2022): 409–440. Web.

Kepuska, Veton, and Gamal Bohouta. "Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)." 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, 2018. 99–103. Web.

Lawson, Luke, Jason Beaman, and Michael Mathews. "Within Clinic Reliability and Usability of a Voice-Based Amazon Alexa Administration of the General Anxiety Disorder 7 (GAD 7)." Journal of medical systems 48.1 (2024): 70-. Web.

Jones, Valerie K et al. "Reducing Loneliness and Improving Social Support among Older Adults through Different Modalities of Personal Voice Assistants." Geriatrics (Basel) 9.2 (2024): 22-. Web.

Platt, K. H. (2024, August 27). SpaceX, Blue Origin Vie to expand Mars-Earth Interplanetary internet. Forbes. Accessed October 24, 2024. https://www.forbes.com/sites/kevinholdenplatt/2024/05/14/spacex-blue-origin-vie-to-expand-mars-earth-interplanetary-internet/

Burleigh, S., Cerf, V., Durst, R., Fall, K., Hooke, A., Scott, K., & Weiss, H. (2002, October). THE INTERPLANETARY INTERNET: A COMMUNICATIONS INFRASTRUCTURE FOR MARS EXPLORATION. USCB. https://sites.cs.ucsb.edu/~ebelding/courses/284/s06/papers/Cerf_Interplanetary.pdf

Nielsen, J. (2024, February 20). 10 usability heuristics for user interface design. Nielsen Norman Group. Accessed October 24, 2024. https://www.nngroup.com/articles/ten-usability-heuristics/

Bollinger, A., Duda, K., Wheelock, D., Moore, S., & Wood, S. (2024, February 2). NASA Technical Reports Server (NTRS). NASA NTRS. Accessed October 24, 2024. https://ntrs.nasa.gov/api/citations/20230014207/downloads/NSL%20IWS24%20Bollinger%20MCO%20FRev.pdf

Johnson, Kevin B., Wei-Qi Wei, Dilhan Weeraratne, Mark E. Frisse, Karl Misulis, Kyu Rhee, Juan Zhao, and Jane L. Snowdon. "Precision medicine, AI, and the future of personalized health care." Clinical and translational science 14, no. 1 (2021): 86-93.

Marín-Coca, S et al. “Hybrid Energy Storage Systems for High Power Spacecraft Missions.” Journal of physics. Conference series 2716.1 (2024): 12097-. Web.

Goodwill, J., Wilson, C., & MacKinnon, J. (2023, July). Current AI technology in Space. NASA NTRS. Accessed October 24, 2024. https://ntrs.nasa.gov/api/citations/20240001139/downloads/Current%20Technology%20in%20Space%20v4%20Briefing.pdf

GOV.UK. (2024, May 15). Cyber Security Risks to Artificial Intelligence. GOV.UK. Accessed October 24, 2024. https://www.gov.uk/government/publications/research-on-the-cyber-security-of-ai/cyber-security-risks-to-artificial-intelligence#findings-of-the-risk-assessment

Rößler, Jeremias et al. “AI Trustworthiness and Quality.” Artificial Intelligence and Software Testing. United Kingdom: BCS Learning & Development Limited, 2022. Print.

Eisenberg, T. (2024, October 11). Cimon, the AI-powered robot, launches a new era in Space Travel. IBM. Accessed October 24, 2024. https://www.ibm.com/new/announcements/cimon-ai-robot-launches-new-era-space-travel

Airbus. (2020, April 15). Cimon-2 makes its successful debut on the ISS. Airbus. Accessed October 24, 2024. https://www.airbus.com/en/newsroom/press-releases/2020-04-cimon-2-makes-its-successful-debut-on-the-iss

Beres, D., & Bell, J. (2022, April 19). *How does your brain make split second decisions?*. Big Think. https://bigthink.com/neuropsych/split-decision-making-process/

Löckenhoff CE. Age, time, and decision making: from processing speed to global time horizons. Ann N Y Acad Sci. 2011 Oct;1235:44-56. doi: 10.1111/j.1749-6632.2011.06209.x. PMID: 22023567; PMCID: PMC3757097.

NASA. (2023a, September 26). Humans to Mars. NASA. https://www.nasa.gov/humans-in-space/humans-to-mars/#preparing

Dobney W, Mols L, Mistry D, Tabury K, Baselet B and Baatout S (2023) Evaluation of deep space exploration risks and mitigations against radiation and microgravity. Front. Nucl. Med. 3:1225034. doi: 10.3389/fnume.2023.1225034

NASA. (2024, March 27). *Space radiation risks*. NASA. https://www.nasa.gov/hrp/radiation/space-radiation-risks/

NASA. (2024a, March 27). *Human factors and behavioral performance*. NASA. https://www.nasa.gov/hrp/human-factors-and-behavioral-performance/

NASA. (2024a, March 27). *Human factors and behavioral performance*. NASA. https://www.nasa.gov/hrp/research-operations-integration/

January 11, 2024 Thursday). HERA Mission: NASA's 45-Day Mars Simulation to Study Human Responses. Space Daily. https://advance-lexis-com.ezproxy2.library.colostate.edu/api/document?collection=news&id=urn:contentItem:6B2Y-XPY1-DYJG-N003-00000-00&context=1516831.

Nasrini J, Hermosillo E, Dinges DF, Moore TM, Gur RC and Basner M (2020) Cognitive Performance During Confinement and Sleep Restriction in NASA's Human Exploration Research Analog (HERA). Front. Physiol. 11:394. doi: 10.3389/fphys.2020.00394

Larson, Lindsay et al. “Team Performance in Space Crews: Houston, We Have a Teamwork Problem.” *Acta astronautica* 161 (2019): 108–114. Web.

Anglin KM, Kring JP. *Lessons from a space analog on adaptation for long-duration exploration missions*. Aerosp Med Hum Perform. 2016; 87(4):406–410.

NASA. (2024a, January 19). *Hera*. NASA. https://www.nasa.gov/mission/hera/

Bardan, R. (2024, February 16). *Martians wanted: NASA opens call for simulated yearlong Mars mission*. NASA. https://www.nasa.gov/news-release/martians-wanted-nasa-opens-call-for-simulated-yearlong-mars-mission/

Donaldson, A. (2024, March 20). *NASA shares progress toward early Artemis Moon missions with crew*. NASA. https://www.nasa.gov/news-release/nasa-shares-progress-toward-early-artemis-moon-missions-with-crew/

Systematic reuse and platforming: Application examples for enhancing reuse with model-based systems engineering methods in space systems development. https://journals-sagepub-com.ezproxy2.library.colostate.edu/doi/epub/10.1177/1063293X17736358

Introduction to Space Science, 2021, p.65-79. https://link-springer-com.ezproxy2.library.colostate.edu/content/pdf/10.1007/978-981-16-5751-1_5?pdf=chapter%20toc

Day, J. C., Ingham, M. D., Murray, R. M., Reder, L. J., & Williams, B. C. (2015, April). Engineering resilient space systems - MIT. http://mers-papers.csail.mit.edu/Journal/2015/Insight Day Resilient/Engineering Resilient Space Systems.pdf

NASA. (2023, October 2). International Space Station. NASA. https://www.nasa.gov/reference/international-space-station/

Dai Z, Lei X, Yang C, Zhao L, Lu L, Li Y. Systematic biomedical research of the NASA Twins Study facilitates the hazard risk assessment of long-term spaceflight missions. Protein Cell. 2019 Sep;10(9):628-630. doi: 10.1007/s13238-019-0628-x. PMID: 31111442; PMCID: PMC6711946.

Francine E. Garrett-Bakelman et al. ,The NASA Twins Study: A multidimensional analysis of a year-long human spaceflight.Science364,eaau8650(2019).DOI:10.1126/science.aau8650

NASA. (2024d, April 10). Institutional Review Board (IRB). NASA. https://www.nasa.gov/institutional-review-board/

Landon, L. B., Slack, K. J., & Salas, E. (2024). Psychology and human performance in space programs. extreme application. CRC Press.

Rienecker, Kira & Grue, Katherine & Paladini, Maria Serena & Frias, Elma & Frattini, Valentina & Borlongan, Mia & Chou, Austin & Torres Espín, Abel & Krukowski, Karen & Ferguson, Adam & Rosi, Susanna. (2023). Combined space stressors induce independent behavioral deficits predicted by early peripheral blood monocytes. Scientific Reports. 13. 10.1038/s41598-023-28508-0.

Hupfeld KE, McGregor HR, Reuter-Lorenz PA, Seidler RD. Microgravity effects on the human brain and behavior: Dysfunction and adaptive plasticity. Neurosci Biobehav Rev. 2021 Mar;122:176-189. doi: 10.1016/j.neubiorev.2020.11.017. Epub 2021 Jan 14. PMID: 33454290; PMCID: PMC9650717.

Rappaport, M. B., & Corbally, C. J. (2022). Neuroplasticity as a Foundation for Decision-Making in Space. NeuroSci, 3(3), 457-475. https://doi.org/10.3390/neurosci3030033

Welch, R. B., Hoover, M., & Southward, E. F. (2009). Cognitive performance during prismatic displacement as a partial analogue of “space fog.” Aviation, Space, and Environmental Medicine, 80(9), 771–780. https://doi.org/10.3357/ASEM.2415.2009

Holden, K., Munson, B., Stephenson, J. (2021). Monitoring Human Performance on Future Deep Space Missions. In: Ayaz, H., Asgher, U., Paletta, L. (eds) Advances in Neuroergonomics and Cognitive Engineering. AHFE 2021. Lecture Notes in Networks and Systems, vol 259. Springer, Cham. https://doi.org/10.1007/978-3-030-80285-1_6

Marquez, J. J., Landon, L. B., & Salas, E. (2023). The Next Giant Leap for Space Human Factors: The Opportunities. *Human factors*, *65*(6), 1279–1288. https://doi.org/10.1177/00187208231174955

Matar, M., Rehm, H., Austin, C., Gokoglu, S., McIntyre, L., & Myers, J. (2024, February 16). *Strategies for Quantifying Human Space Flight Performance in the Crew Health and Performance System* (Paper No. 20240001176). Presented at the Human Research Program Investigators’ Workshop, Galveston, TX. National Aeronautics and Space Administration. https://ntrs.nasa.gov/api/citations/20240001176/downloads/IWS2024_Matar_Final.pdf

Hamlin, T. L. (2010, October 26). *The importance of human reliability analysis in human space flight: Understanding the risks* (Report No. JSC-CN-21806) [Presentation]. National Aeronautics and Space Administration. https://ntrs.nasa.gov/api/citations/20100036673/downloads/20100036673.pdf?utm_source=chatgpt.com

Lai, J.-S., & Selby, R. (2016). Human systems integration: Process to help minimize human errors, a systems engineering perspective for human space exploration missions. *REACH – Reviews in Human Space Exploration, 2*, 8–23. https://doi.org/10.1016/j.reach.2016.11.003

Kanas, N., & Manzey, D. (2008). *Space Psychology and Psychiatry* (2nd ed.). Springer.

Palinkas, L. A., & Suedfeld, P. (2021). Psychosocial issues in isolated and confined extreme environments. *Neuroscience and biobehavioral reviews*, *126*, 413–429. https://doi.org/10.1016/j.neubiorev.2021.03.032

Rosenberg, J. D., Jannasch, A., Binsted, K., & Landry, S. (2022). Biobehavioral and psychosocial stress changes during three 8–12 month spaceflight analog missions with Mars-like conditions of isolation and confinement. *Frontiers in Physiology, 13*. https://doi.org/10.3389/fphys.2022.898841

Flynn-Evans, E., Barger, L., Kubey, A. *et al.* Circadian misalignment affects sleep and medication use before and during spaceflight. *npj Microgravity* **2**, 15019 (2016). https://doi.org/10.1038/npjmgrav.2015.19

Sandal, Gro. (2001). Crew Tension during a Space Station Simulation. Environment and behavior. 33. 134-50. 10.1177/00139160121972918.

Frank, J. D. (2019). *Artificial intelligence: Powering human exploration of the Moon and Mars* (NASA/TM–2019–220371). NASA Ames Research Center. https://ntrs.nasa.gov/api/citations/20190032627/downloads/20190032627.pdf

Straub, J. (2011, July). *A review of spacecraft AI control systems* [Conference paper]. 15th World Multi-Conference on Systemics, Cybernetics and Informatics (SCI 2011), Orlando, FL. https://www.iiis.org/CDs2011/CD2011SCI/SCI_2011/PapersPdf/SA517LI.pdf

Li, L., & Li, Y. (2011). *A review of spacecraft AI control systems*. In *Proceedings of the 15th World Multi-Conference on Systemics, Cybernetics and Informatics (SCI 2011)*. International Institute of Informatics and Systemics. https://www.iiis.org/CDs2011/CD2011SCI/SCI_2011/PapersPdf/SA517LI.pdf

P. A. Oche, G. A. Ewa and N. Ibekwe, "Applications and Challenges of Artificial Intelligence in Space Missions," in IEEE Access, vol. 12, pp. 44481-44509, 2024, doi: 10.1109/ACCESS.2021.3132500.

Appendix A: Survey Questions

Below is a printout of the survey given to the participants.

Thank you for participating in this survey. My name is Stephanie Anderson and I am conducting this research as part of my PhD Dissertation in Systems Engineering at Colorado State University. The purpose of this study is to understand your preferences and perceptions regarding the integration of artificial intelligence (AI) in spacecraft during long-duration spaceflights. Your responses will help guide the design of AI systems for future space missions.

This survey consists of three sections: Background Information, Discrete Choice Tasks, and Additional Feedback. It should take approximately 20–30 minutes to complete. All responses are anonymous, will be kept strictly confidential, and will be used solely for academic purposes. No personal identifiers will be included in any reports or publications resulting from this study.

Should you have any questions or require further information, please feel free to contact me at Stephanie.Anderson@colostate.edu.

1. By selecting 'Yes' below and continuing, you are voluntarily agreeing to participate in this research study. You may withdraw from the survey at any time.

◯ Yes

◯ No

Section 1: Background Information

Please check the box that best describes you for each question:

2. Experience Level:

- ☐ Astronaut with spaceflight experience
- ☐ Astronaut with no spaceflight experience
- ☐ Analog astronaut (simulated mission experience)
- ☐ Other (please specify)

3. Familiarity with AI Systems:

- ☐ Very familiar (have used AI extensively)
- ☐ Somewhat familiar (have a basic understanding of AI)
- ☐ Not familiar (little to no experience with AI)
- ☐ Other (please specify)

4. Preferred Role for AI in Decision-Making:

- ☐ Fully autonomous (AI makes all decisions)
- ☐ Collaborative (shared decision-making with AI)
- ☐ Human-driven (humans make all decisions with AI as a tool)

5. Previous Experience with AI or Automation in High-Stakes Settings:

- ☐ Yes
- ☐ No
- ☐ If yes, please describe

Section 2: Discrete Choice Tasks

Next are several scenarios involving AI systems in a spacecraft. Each scenario varies based on conditions in which AI would be used. Please review each scenario and select the options you would most prefer.

Scenario #1: Life Support

Maintaining Environmental Control and Life Support Systems (ECLSS). Setting optimal temperature and maintaining cabin pressure and humidity levels. Allocating rationing of food and water supplies to support mission duration and crew health.

6. At what level should the AI operate in this scenario?

- ◯ Advisor (suggests actions)
- ◯ Assistant (executes commands under supervision)
- ◯ Autonomous Controller (makes independent decisions)

7. What is the level of reliability of the AI needed for this scenario?

- ◯ High (99% and higher error-free)
- ◯ Medium (95% - 98% error-free)
- ◯ Low (90% or lower error-free)

8. What level of transparency should the AI have in this scenario?

- ◯ Fully explainable decisions
- ◯ Partially explainable decisions
- ◯ Opaque decisions (black-box AI)

9. What level of human override capability should there be in this scenario?

- ◯ Always possible
- ◯ Sometimes possible
- ◯ Not possible (critical operations)

10. How confident are you in your above choices?

- ◯ Extremely confident
- ◯ Very confident
- ◯ Somewhat confident
- ◯ Not so confident
- ◯ Not at all confident

Section 2: Discrete Choice Tasks (cont.)

Scenario #2: Routine Tasks

Performing routine tasks such as:
- Propellant level checks
- Trajectory accuracy checks
- Equipment functionality checks
- Visually inspecting spacecraft interior and exterior for potential issues

11. At what level should the AI operate in this scenario?

- ◯ Advisor (suggests actions)
- ◯ Assistant (executes commands under supervision)
- ◯ Autonomous Controller (makes independent decisions)

12. What is the level of reliability of the AI needed for this scenario?

- ◯ High (99% and higher error-free)
- ◯ Medium (95% - 98% error-free)
- ◯ Low (90% and below error-free)

13. What level of transparency should the AI have in this scenario?

- ◯ Fully explainable decisions
- ◯ Partially explainable decisions
- ◯ Opaque decisions (black-box AI)

14. What level of human override capability should there be in this scenario?

- Always possible
- Sometimes possible
- Not possible (critical operations)

15. How confident are you in your above choices?

- Extremely confident
- Very confident
- Somewhat confident
- Not so confident
- Not at all confident

Section 2: Discrete Choice Tasks (cont.)

Scenario #3: Emergency (non-critical)

CO2 levels begin to rise in the spacecraft due to a malfunction with the primary CO2 scrubber. Backup scrubber is activated but the primary scrubber needs troubleshooting, evaluation, and potential fixes. AI could be used for the troubleshooting/evaluation/fix.

16. At what level should the AI operate in this scenario?

- Advisor (suggests actions)
- Assistant (executes commands under supervision)
- Autonomous Controller (makes independent decisions)

17. What is the level of reliability of the AI needed for this scenario?

- High (99% and higher error-free)
- Medium (95% - 98% error-free)
- Low (90% and below error-free)

18. What level of transparency should the AI have in this scenario?

- Fully explainable decisions
- Partially explainable decisions
- Opaque decisions (black-box AI)

19. What level of human override capability should there be in this scenario?

- ◯ Always possible
- ◯ Sometimes possible
- ◯ Not possible (critical operations)

20. How confident are you in your above choices?

- ◯ Extremely confident
- ◯ Very confident
- ◯ Somewhat confident
- ◯ Not so confident
- ◯ Not at all confident

Section 2: Discrete Choice Tasks (cont.)

Scenario #4: Emergency (Mission Critical)

During spaceflight, alarms sound indicating a sudden drop in cabin pressure. A micrometeorite has struck the spacecraft, puncturing a small but dangerous hole in the hull. Air is rapidly escaping and oxygen levels are dropping. If not addressed immediately, crew risks losing breathable air, leading to asphyxiation or structural failure of the spacecraft.

21. At what level should the AI operate in this scenario?

- ◯ Advisor (suggests actions)
- ◯ Assistant (executes commands under supervision)
- ◯ Autonomous Controller (makes independent decisions)

22. What is the level of reliability of the AI needed for this scenario?

- ◯ High (99% and higher error-free)
- ◯ Medium (95% - 98% error-free)
- ◯ Low (90% and below error-free)

23. What level of transparency should the AI have in this scenario?

- ◯ Fully explainable decisions
- ◯ Partially explainable decisions
- ◯ Opaque decisions (black-box AI)

24. What level of human override capability should there be in this scenario?

- Always possible
- Sometimes possible
- Not possible (critical operations)

25. How confident are you in your above choices?

- Extremely confident
- Very confident
- Somewhat confident
- Not so confident
- Not at all confident

Section 3: Additional Questions

Please provide feedback for the following questions related to crewed spaceflight.

26. How much do you trust AI systems in high-stakes environments?

- Fully trust
- Somewhat trust
- Neutral
- Somewhat distrust
- Fully distrust

27. How comfortable are you with AI making critical decisions without human involvement?

- Very comfortable
- Somewhat comfortable
- Neutral
- Somewhat uncomfortable
- Very uncomfortable

28. What level of risk would you accept for an AI failure in life support systems?

- Minimal
- Moderate
- Severe

29. What level of risk would you accept for an AI failure in navigation and docking?

- Minimal
- Moderate
- Severe

30. What level of risk would you accept for an AI failure in an emergency response situation?

- Minimal
- Moderate
- Severe

31. What level of risk would you accept for an AI failure in a scientific experiment?

- Minimal
- Moderate
- Severe

32. How important is it for an AI system to explain its decisions?

- Extremely important
- Very important
- Somewhat important
- Not so important
- Not at all important

33. Would you want to take part in the development and training of the AI system being used on a spacecraft you will fly on?

- Yes
- No

Section 4: Final Thoughts

Thank you for participating in this survey. Your responses will help shape how AI is used in future spaceflight endeavors.

34. Please add any additional comments you would like to share regarding AI integration in spacecraft for long duration missions.

Appendix B: Additional Feedback from Open-Ended Question at End of Survey

Below is a table with the gathered open-ended responses. Each cell represents an individual's response.

Table 3. Open-Ended Survey Responses

I think integrating an AI model depends very heavily on having crew work with it and train with it in a variety of scenarios so they can understand how it functions, and that they can have confidence in robust testing of it. I think AI is an incredibly useful tool, but it is critical that the actions AI can take are bounded and overridable to enable human operators to correct any unanticipated behaviors.
Any automation or autonomous systems are ultimately programmed/created by humans; therefore, errors or shortsighted programming are guaranteed to occur. Automation is great for crew relief; the crew just needs the insight to be able to understand what automation is doing or has done in order to verify automation actions (building trust in the system) or take over from automation at any time.
While speed is often essential for reaction to an emergency, once the systems are safe, it's also important to understand how you got there.
If we start using AI for our operations, I would like to understand it's development process as well. Ex. AI logic, training process etc.
Typically, AI systems are constantly evolving, based on human feedback. There are, however, circumstances where AI needs to be "frozen", and I think spaceflight should be one of them. That way the flight version would remain identical to the ground version, allowing scenario checks and simulations.
Current AI is prone to serious errors. It will confidently tell you lies, and when questioned it will lie about its lying. A crewmate with that behavior would be deassigned from the flight early in training and would never be allowed to fly in space. As with a human crewmate, AI must earn the absolute trust of people whose lives will depend on its judgment and actions. A barrier to achieving this is conscientiousness. A socially normal human who makes a mistake generally feels bad about it and accepts that negative

emotion as impetus to improve in the future. An AI model might mimic that response but will never actually experience it. Hence it can never be trusted to learn and improve.
I approached and answered these questions more as an experienced engineer than as an analog crewmember, because these are generally engineering questions (e.g. they represent desires and decisions made during the design process) instead of operations questions. In this context, I'm generally using my risk mitigation and FMECA tools to compare AI reliability to the reliability required of typical flight hardware and software components. One critical aspect of flight hardware design is that it needs to be verified against its functional requirements and validated in the system it is integrated into; this design approach guided many of my answers.
I believe AI is a great tool as long as it doesn't choose to solve a problem in way that damages the passengers safety. For minor maintenance tasks it is great at taking that off of your plate and in emergencies providing a quick update on what is going on, but h wouldn't trust it implicitly or over the experience of my crew
I think an important part of the context of AI use is what level of human support/oversight/workflow changes are considered with AI integration. I also think it is underutilized for routine tasks but also part of it is how it confirms what it did/how it did it, etc. The human interface for these technologies is a critical part of acceptance, viability, utility, etc. Thus, the way AI is valued must be done within the context of the human interface.
Explainability is important, but not always even theoretically possible with the best performing systems. I would be more comfortable with an AI system that could be shown to beat human performance baselines, even if it could not fully explain to me reasoning behind every decision.
To copy from IBM: A computer cannot be held accountable, therefore a computer cannot make a decision.
Progress in AI has been substantial in the last several years; however, I believe I have a healthy distrust in relying on AI to solve problems independently. If the lives of my crew are at risk, I would want to make sure the crew uses the AI as a tool, not a crutch. I do recognize that in emergency situations or in very complex environments having an AI assistant would be beneficial, assuming that it can fully explain its suggestions and how

it worked through the problem at hand. At the end of the day, it is the human crew in the spacecraft who are at risk. I believe the crew should have control over any and all systems on board (including the AI).
I believe having AI integrated into spacecraft and operations for long duration missions is critical; however, I think the greatest gap to bridge is AI automation of addressing certain criticalities and minimizing error rate to near zero.
Being an operating engineer on shipboard systems on 20-40 year old systems, there is absolutely no AI integration. Therefore, human element is critical for understanding, operating, troubleshooting, and fixing systems. As new systems become available, we need to ensure that the human engineer remains in the loop, so as not to devolve knowledge and experience base. AI does have a role of facilitating routine operations, but the human engineer must remain in the loop.
Dear Stephanie, Thank you for considering me for your research. Wishing you the best of luck with your study and dissertation. Warm regards, Pietro Di Tillio
I believe trust in AI can be built. If astronauts see how AI makes decision in analog situations they can have better trust. But the AI is only as good as it has been trained - that's the hesitancy with fully trusting it at this point. But there are many examples of fully autonomous new technologies that initially people were very nervous about but now there is full trust. In fact, they are safer than if humans intervened. But that takes time. I think that will be the same here. Trust with AI should be earned slowly with time and experience. (I would be interested to see where this study goes and how it impacts training and procedure in the future.)
We still do not understand AI decision making so human override capabilities on a spacecraft leaving near Earth orbit are critical to having confidence in the proposed systems. Plus, I don't think we can fully anticipate what scenarios the first humans traveling to Mars might encounter.
There's a lot of nuance and assumptions baked into my previous answers. For a future survey, I recommend adding an individual comment box for each question if possible. In general, my answers for all of the risk-based questions depend on the worst case thing that could happen for an improper action for the AI system - e.g. is there risk of permanent equipment damage or human injury/death from an AI system's actions on a

timeline faster than a human would reasonably be able to prevent it? If manual intervention is expected (or even possible) in some point of the operation loop of the given thing in time to catch an error, whether it's docking, ECLSS, or anything else, we should have higher risk thresholds for failure. I recently left Blue Origin to join a frontier AI lab (Anthropic) and would be happy to explain my answers or other thoughts on the intersection of AI and human spaceflight if useful. Thanks for the chance to engage on this! -Barret barret@anthropic.com
Use and trust of AI and task automation is something I look to calibrate with over time. Answers can look different with time and experience, both in nominal and off nominal conditions.
Here are several thoughts I believe are important: (1) Human readable logging. Rather than a scale of "Explainable" to "Opaque," an AI user, I'd like options of logged explanations from "Highly Detailed" to "Generic." And these explanations could also be categorized either (a) in vivo calculated explanation, where the model is explaining how it calculated its response; or (b) an ex-post facto explanation, where the model, after a logged action and when it has processing available, makes an assessment as to how it made a decision based on logged results. As a subset of functionality, the logging output in critical, time-sensitive decisions could be more minimal while tuned to be assessed and explained later. This might limit the calculation time during operations while still allowing for assessment and explanation later. (2) Blind spot assessments. If I was working to develop an AI for use on an LTSE mission, I'd like the model to know what my strengths are, say my CV and a set of tailored questions. In for science, management, and operations, I'd like to have the model help me implement strengths and, more importantly, point out my intellectual and operational blind spots before making decisions. Where the model calculates that there are solutions outside of my knowledge base, I'd like the model to provide thought-provoking responses a la "I am an expert in X subject who is trying to solve a problem in Y subject. What are areas where X subject is helpful, not helpful, in solving a problem in Y subject. Finally, what are areas in Y subject that X subject does not cover or contradict?" (3) Integrated training tasks. Depending on what a model will be used for, I'd like to have an AI already integrated into my work. Preferably, using an AI wouldn't be anything different than my normal work routines. However, particularly when it comes to operational, critical, or emergency tasks, I'd like to have clear protocols for engaging with an AI system including clear failsafe for "analog" operations without an AI (What if the micrometeorite pierces the server and all of its backups?). This would also mean designing scientific studies to have non-AI fallback strategies as well. I know this is obnoxious but I see Hal's malfunction in 2001: A Space Odyssey. The description of Hal's error is actually quite realistic, instructive, and prescient. If an AI model is trained improperly or given hidden or conflicting system prompts, the model's output, responses, operations, or possibly anything the model is connected to, might be tainted. Two recent examples are demonstrative of the peril:

First, xAI was recently reported to have issues with the system prompts in Grok which generated less than optimal output. Second, OpenAI was reported to have recently rolled back a model due to the style of the model's interactions with users. In the potential contexts which might be extant during an LSTE mission, these issues can be disastrous.

Appendix C: NASA IRB Study Approval Letter

National Aeronautics and Space Administration

Office of the Chief Health and Medical Officer (OCHMO)
Mary W. Jackson NASA Headquarters | Washington D.C.

Notification of Approval

22May2025

TO: Stephanie Anderson
7135692745
stephanie.s.anderson@boeing.com

FROM: Marisa Covington, Ph.D., CIP
Chair, NASA Institutional Review Board

TITLE: Behavioral Modeling Experiment for Integrating Artificial Intelligence on Spacecraft for Long-Duration Missions

Study eIRB Number	STUDY00000855
Type of Review:	Initial Study
Method of Review:	**Exempt Review – Category (2)(i)**
IRB Disposition:	Approved
Date of IRB Determination:	22May2025
Effective Date:	22May2025
Risk Level:	No greater than minimal risk
FWA Number:	00019876

The above referenced proposal has been reviewed and approved by the NASA IRB in accordance with ethical standards and the requirements of the Code of Federal Regulations on the Protection of Human Subjects (NASA 14 CFR 1230 and if applicable, FDA 21 CFR 50 and 56).

IRB approval is valid effective 22May2025.

To maintain compliance with federal regulations and NASA policy:

- All modifications, revisions, or changes to this protocol must be submitted to the NASA IRB for review and approval prior to implementation.
    - Changes, including responses to committee requests do not extend the initial review and/or protocol renewal date. There is no grace period beyond the approval date.

- Conflict of Interest (COI) disclosure forms must be updated annually for all study personnel.

- PI are responsible for maintaining current CITI training documentation for all study personnel. This documentation may be updated in the eIRB on an annual basis at the time of Continuing Review.

- Once all research activities have been completed and data permanently deidentified (or deposited with LSDA) the PI should request closure of IRB oversight for the project within the eIRB.

- Adverse events or unexpected problems (UPIRSOs) resulting from this study to the NASA IRB, sponsor/funding source, and the Safety Office (if applicable).

The Principal Investigator is responsible for following all pertinent ethical and legal guidelines as well as NASA policies and directives.

IRB approval means that the requirements of federal regulations and NASA policies governing human subject research have been met. **Approval from other entities may also be needed. NASA IRB approval in no way implies or guarantees that permission from these other entities will be granted.**

Sincerely,

Marisa Covington, Ph.D., CIP
Chair, NASA IRB