# Advancing Intelligent Personal Assistants for Human Spaceflight


**Leonie Bensch**[1]
(MIT) Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

**Oliver Bensch**
Technical University of Munich (TUM), Germany

**Tommy Nilsson**
European Space Agency (ESA), Cologne, Germany



## Abstract

The Artemis program and upcoming missions to Mars mark a new era of human space exploration that will require new tools to support astronaut autonomy in the absence of real-time communication with Earth. This paper investigates the role of voice-based intelligent personal assistants (IPAs) in future crewed space missions. Through semi-structured interviews with astronauts (n=3) and spaceflight experts (n=12), we identify key user-centered design requirements for IPAs in this uniquely constrained and safety-critical environment. Our thematic analysis reveals core requirements for flexibility, reliability, offline capability, and multimodal interaction. Drawing on these findings, we outline design guidelines for next-generation IPAs and discuss how technologies such as retrieval-augmented generation (RAG), knowledge graphs, and augmented reality should be combined to support flexible, reliable, and multimodal IPAs for future human spaceflight missions.



**2012 ACM Subject Classification** Human-centered computing → Empirical studies in HCI; Computing methodologies → Intelligent agents

**Keywords and phrases** Conversational Assistant, Intelligent Personal Assistant, Artificial Intelligence, Astronaut, Human Spaceflight, Generative Pre-Trained Transformer (GPT), Retrieval Augmented Generation (RAG), Knowledge Graphs, Augmented Reality, Voice Assistant, Long Duration Spaceflight

**Digital Object Identifier** 10.4230/OASIcs.SpaceCHI.2025.18


## 1 Introduction

Amidst a surging interest in human space exploration, ambitious programmes aiming to establish a sustained human presence on the Moon are now underway, with crewed missions to Mars projected to follow soon after [43]. The nature of these missions represents a shift not only in distance and duration, but also in the complexity and level of autonomy required from astronauts.

Historically, astronauts operating in Low Earth Orbit or on the lunar surface relied on relatively rudimentary tools such as printed checklists while benefiting from uninterrupted, real-time supervision by ground control [33, 31, 18, 28, 21]. In contrast, future missions will involve operations of unprecedented complexity, including the assembly of surface infrastructure and in-situ resource utilisation, placing significant cognitive and physical demands on the crew.

These challenges are further increased by communication constraints. The Artemis lunar programme, for example, is targeting the Moon's south pole, an area prone to recurring blackouts due to topographical shadowing and orbital libration, making reliable contact with

[1] Corresponding author



Advancing Human-Computer Interaction for Space Exploration (SpaceCHI 2025).
Editors: Leonie Bensch, Tommy Nilsson, Martin Nisser, Pat Pataranutaporn, Albrecht Schmidt, and Valentina Sumini; Article No. 18; pp. 18:1–18:18
OpenAccess Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Earth difficult [48, 19, 8, 17, 23]. Similarly, during future Mars missions, astronauts will experience latencies of up to 42 minutes in round-trip communication with Earth [24]. In such contexts, autonomy is not a supplementary feature but a mission-critical necessity.

To support this autonomy, voice-based intelligent personal assistants (IPAs) are attracting growing attention in the space domain. These systems promise hands-free, naturalistic interaction and could serve as intermediaries between astronauts and a wide range of operational systems. Thanks to advances in artificial intelligence, IPAs may be able to provide assistance even in dynamic or time-sensitive scenarios.

Prototypes, such as METIS [6], which supports ground control, CIMON tested on the International Space Station (ISS) [13] or CORE [5], designed to aid astronauts during long-term missions, point in this direction. Yet, their development has largely prioritised technological feasibility, with limited focus on their human-centered aspects, such as usability and other user requirements in the context of future human spaceflight [6].

In this work, we examine IPAs not merely as retrieval interfaces but as collaborative systems supporting astronaut autonomy under constrained and safety-critical conditions. The study is guided by the following research question: *What are the core user-centered requirements for IPAs in future space missions beyond Low Earth Orbit?*. Drawing on semi-structured interviews with astronauts (n=3) and domain experts (n=12), we explore the core requirements for IPAs in future space missions. Our findings point toward the need for flexible, reliable, and locally deployable (offline) multimodal systems capable of adapting to evolving operational demands.

In addressing the central research question, we present the following contributions: 1) A qualitative analysis of IPA requirements derived from interviews with astronauts and space experts. 2) The identification of four central themes: flexibility, reliability, offline availability, and multimodal feedback, each contextualised within specific operational scenarios, and lastly 3) initial directions on how to target each theme from a technical standpoint, using knowledge graphs (KGs), Retrieval Augmented Generation (RAG), and Augmented Reality (AR) cues.

## 2 Related Work

Over the past decade, IPAs have become increasingly relevant to space exploration, particularly as voice interfaces mature and onboard computing capabilities improve. Early deployments in low Earth orbit, such as CIMON (see Figure 1), have laid the groundwork for their potential use in future missions.

CIMON, launched to the International Space Station in 2018, was one of the first artificial intelligence-powered assistants tested in space [13]. Based on IBM Watson, it supported basic voice interaction and experiment assistance but relied on internet connectivity and could not function autonomously. Its successor, CIMON 2, included minor improvements such as facial recognition and sentiment (i.e., analysis of emotional tone) detection, but remained rooted in narrow artificial intelligence principles, offering limited adaptability or learning capabilities [38].

More recently, Amazon's Alexa was integrated into the Artemis I mission as part of a technology demonstration, signaling commercial interest in extending voice assistants to the space domain [1]. Similarly, the AI4U assistant, developed using reinforcement learning, is being tested under Mars analogue conditions with the aim of future International Space Station deployment [42, 34, 45]. While promising, AI4U requires extensive pre-training and has yet to be validated in real operational settings.

**Figure 1** CIMON, an early example of an intelligent personal assistant deployed aboard the International Space Station. It served as a proof of concept for voice interaction in microgravity but was limited by its reliance on narrow AI and Earth-based processing.

Other approaches, such as METIS [6] and CORE [5], integrate RAG and KG technologies to enable more structured and reliable information delivery. These systems represent a departure from earlier rule-based assistants and reflect growing interest in generative models. However, existing work has largely focused on feasibility rather than usability. Questions of trust, explainability, and interaction under communication constraints remain underexplored.

Within the human-computer interaction community, there is a growing recognition of the need for more user-centered approaches to artificial intelligence design, particularly in safety-critical or resource-constrained environments. Prior work in aviation, healthcare, and remote operations has shown that trust, explainability, and multimodal feedback are essential to support effective human-AI collaboration [27]. However, these findings have yet to be fully applied to the spaceflight context, where procedural rigidity, environmental constraints, and isolation introduce additional design challenges.

With this study, we aim to help lay the groundwork for closing this gap. Through interviews with astronauts and domain experts, we examine IPA design not only as a technical challenge but as a human-centered one. In doing so, we contribute a perspective grounded in astronaut experience and operational realities, offering a foundation for the next generation of IPA systems in human spaceflight.

## 3 Methodology

### 3.1 Participants

For this study on IPAs for future human spaceflight missions to the ISS, Moon, and beyond, we selectively invited experts from a range of disciplines related to space operations and planetary exploration. Participants were recruited via purposive sampling based on their expertise in astronaut operations, training, or space technology development. Invitations were extended through professional networks, institutional affiliations, and referrals. The final sample included 15 domain experts, among them 3 active astronauts with a combined total of approximately 900 days in space and over 20 hours of extravehicular activity (EVA) on the ISS (see Table 1).

In addition to astronauts, the group included engineers with mission planning and crew training experience, instructors supporting astronaut training and EUROCOM operations, and scientists involved in lunar and Martian field studies. All participants reported prior use of conversational agents such as Alexa, Siri, Google Home, or ChatGPT, and several had direct experience with CIMON, the AI assistant deployed on the ISS.

**Table 1** Summary of participant demographics. Areas of expertise are self-reported.

| Role | Gender | Job Title | Area of Expertise |
|---|---|---|---|
| Astronaut 1 | F | Astronaut | Human Spaceflight, EVA |
| Astronaut 2 | M | Astronaut | Human Spaceflight, EVA |
| Astronaut 3 | M | Astronaut | Human Spaceflight, EVA |
| Engineer 1 | M | Aerospace Engineer | Aerospace Engineering, Astronaut Training |
| Engineer 2 | F | Engineer | Human Factors and Ergonomics |
| Engineer 3 | M | VR/AR Software Engineer | VR/AR Software Engineering |
| Engineer 4 | M | In-Space Manufacturing Specialist | In-Space Manufacturing |
| Engineer 5 | M | In-Space Manufacturing Specialist | In-Space Manufacturing |
| Instructor 1 | M | Astronaut Training and Simulations Support | Training and Simulations of EVAs |
| Instructor 2 | M | Astronaut Training and Simulations Support | Training and Simulations of EVAs |
| Instructor 3 | F | Training and EUROCOM Specialist | Training and EUROCOM |
| Scientist 1 | F | Research Fellow | Machine Learning |
| Scientist 2 | M | Lunar Scientist | Lunar Science, LUNA Facility, Aerospace |
| Scientist 3 | M | Physicist | Physics, EUROCOM Crew Support |
| Scientist 4 | M | Space Radiation Shielding Specialist | Space Radiation Shielding |

## 3.2 Procedure

Participants were invited to take part in semi-structured interviews conducted in a quiet room at our lab at the European Space Agency's European Astronaut Centre (ESA-EAC). All participants provided informed consent prior to the interviews. The interviews lasted around 30 minutes. A pre-formulated interview guide was used to ensure consistency across sessions, while still allowing room for follow-up questions and deeper exploration of interesting topics raised by the participants. We observed that thematic saturation was reached by the 13th interview, with no new themes emerging in the final two sessions. The interview questions focused on the following areas:

- *What are the essential requirements that IPAs must fulfill to be used in future space operations?*
- *What are the critical shortcomings of existing IPAs, such as Alexa, Siri, Google Home, and CIMON, that could hinder their effectiveness in space operations?*
- *Do the requirements for IPAs vary between missions to the ISS, the Moon, Mars, or extended-duration spaceflight? If so, how?*

The interview guide was based on previous work in human-computer interaction, IPAs, and extreme environment operations (e.g., spaceflight, military). It was reviewed by two domain experts to ensure relevance and clarity. When participants raised unanticipated points, we asked follow-up questions to explore those aspects in more detail.

All interviews were conducted by a member of the research team familiar with the topic of space systems and human-agent interaction, which helped establish rapport and contextual understanding. Each session was audio-recorded and transcribed verbatim by the research team.

### 3.3 Data Analysis

We conducted a thematic analysis following the six-phase approach described by Braun and Clarke (2006) [9]. This method was chosen for its flexibility and suitability in identifying patterns and meanings across qualitative datasets. The process included familiarization with the data, initial code generation, theme identification, theme review, theme definition and naming, and final report production.

Two researchers independently coded the transcripts using a manual approach (without software assistance). Code themes were initially developed inductively from the data, without relying on any pre-existing coding frame. After independent coding, the researchers met to compare their respective findings, resolve discrepancies through discussion, and consolidate them into a shared coding scheme. This collaborative process served as an informal approach to establishing inter-coder agreement, ensuring that themes were consistently and accurately represented.

Emergent themes were iteratively refined and synthesized into a coherent thematic structure that captured the key insights and challenges regarding the deployment of IPAs in human spaceflight contexts.

## 4 Qualitative Findings

### 4.1 Flexibility

The interviews highlighted the importance of flexibility in IPA systems, emphasizing the need for adaptability across a wide range of tasks and contexts. This included not only practical applications such as checklist completion, suit monitoring, and navigation, but also the ability to modulate behavior based on different scenarios, emergencies versus routine tasks, and user preferences. Participants repeatedly pointed out that IPAs should offer a degree of personalization and responsiveness to unforeseen situations.

**Scientist 4** envisioned a system that supports dynamic adaptation: *"A personalized setup… that is also flexible depending on the activity."*

**Instructor 2** framed adaptability in terms of learning from astronauts' training performance: *"If you know you will operate with someone, you will train with this person, spend some time with them, and adapt. At the moment, we're only adapting the way we're communicating with and treating astronauts. We don't adapt procedures, for example. But procedures could be adapted using AI. During the training it might learn: 'Hey, this crew member is always struggling with this one step.' So when the crew member reaches this step, they could be shown a video of him or her performing it during training. Operationally, that could be very useful."*

A recurring theme was the need for a flexible communication style. Participants wanted to be able to tell the assistant how to behave. **Engineer 4** noted that astronauts may prefer concise instructions: *"Not verbose, or like very quick."*

**Instructor 3** reflected on the possibility of training the AI to adapt emotionally: *"Your Alexa you can also train to be more emotional or more rational. If you keep shouting at your Alexa and cursing at it, it also becomes very...[unfiendly]. But if you say 'Hey, how are you this morning?', then Alexa also starts responding in a more friendly, emotional, human kind of way."*

This personalization extended into entertainment and well-being. **Engineer 5** described the importance of tailoring the assistant's tone and content to the user's mood and context: *"It's super personal. Some people really love it if they get jokes every five minutes. Others just want military-style instructions, without being bothered by any side-talk. Some might find it useful if their heart rate is up and somebody plays classical music. But others might think, 'Stop this and let me work.' That's something the AI could do for you or offer you."*

While some participants were open to emotionally adaptive responses, many expressed discomfort with AI systems that mimic human behavior too closely. **Scientist 1** and **Engineer 4** raised concerns about systems attempting to simulate personalities or emotional intelligence without sufficient realism.

**Scientist 1** shared: *"I always find it super creepy when you try to make something more human-like. You know, these kinds of things where it looks too close to being real, but the technology is not good enough. I think because AI is the only thing that, no matter what, is trained on a subset of data. Whereas a person can take in all the information, and it's also culture. It's not flexible."*

**Engineer 4** recounted his experience simulating AI characters: *"My friend and I pretended to give our character to ChatGPT, but then you could always tell it was simulated. It wasn't real. I would find that unnerving. Unless the AI is extremely good at fooling me or really consistent in its answers, I would rather not have a personality. It needs to be consistent, or it takes away the immersion."*

**Engineer 5** echoed this sentiment, preferring machine-like behavior over human mimicry: *"It would probably be a computer-style voice rather than pretending to be human. Right now, it is pretending to be human with feelings and emotions. If the AI knows you're stressed, it will change its voice level like a human would. That feels wrong to me. It reads my heart rate and talks to me calmly. But if I'm relaxed and it knows I'm an hour behind schedule, it talks differently again. That might feel wrong, although maybe with time, I would get used to it."*

Although a few participants valued social and emotional cues in voice interaction, the majority favored a more neutral, logical assistant, especially in high-stakes or emergency contexts. **Scientist 1** summarized this view: *"If you're actually in a life-threatening situation, you want precise answers. It's okay not to be joking around too much."*

Overall, participants emphasized that flexibility should not only address task support but also adapt to user preferences, emotional context, and operational demands. However, this flexibility should be implemented with care to avoid overstepping into unrealistic or unsettling human imitation.

## 4.2 Reliability

The importance of IPA's reliability consistently surfaced as another key theme in the interview study. Drawing from their first-hand experience with AI deployments such as CIMON aboard the ISS and commercial systems like Alexa or ChatGPT, many participants expressed skepticism about the current technological maturity of AI assistants in spaceflight contexts.

**Engineer 1:** *"Everything we've tried so far worked very poorly. Technologically, some of the solutions that we've used were of the classical type, where you had to train the system to recognize your voice. And even then, the performance was not great, because the system itself was really limited in computational power... It was consistently terrible. [...] If you have a solution for operational use, it needs to work 100% of the time."*

This sentiment was echoed by another participant with engineering responsibilities for in-mission operations:

**Engineer 2:** *"In space operations, things have to work. Not six times out of ten, not seven times out of ten – but ten times out of ten. If it works, then sure, but it needs to work reliably and consistently every single time."*

Several participants emphasized that current AI assistants often fail to meet those standards, citing issues such as long response times, unexpected behaviors, and inaccurate outputs. These shortcomings were viewed as particularly problematic in safety-critical settings such as EVAs.

**Instructor 1:** *"The problem is, for the time being, if you're completely relying on such an artificial intelligence assistant, you still have some unexpected behaviours. [...] I'm not sure if I would put an AI assistant immediately in a [safety] critical environment, like an EVA or Moon exploration. I would rather put it in a support function, until it's really reliable in the response."*

**Scientist 1** added that the narrow training scope of most IPAs makes them ill-suited for complex, dynamic environments, suggesting that human crew members are still more capable of holistic judgment.

There was also a strong preference for human communication over AI in situations where ground contact remains viable, such as on the ISS. Participants highlighted the value of interpersonal trust, shared context, and linguistic nuance, all of which are difficult to replicate in human-AI interaction.

**Instructor 2:** *"For an ISS EVA we have immediate direct communication. I'm sure the crew would always want to talk to their capcom, because they know them, they've trained together. [...] Why would you not talk to someone on the ground and go to the AI? There would just be no advantage, it would only introduce a risk of miscommunication."*

Taken together, these insights suggest that while IPAs may play a supportive role in routine or non-critical tasks, current limitations in their reliability and predictability preclude their use as a substitute for trusted human interaction, especially in mission-critical scenarios.

## 4.3 Offline Availability

The necessity for astronauts to operate independently from terrestrial mission control centers likewise emerged as a crucial requirement across interviews. Participants consistently emphasized that, in the absence of real-time communication, particularly during missions to the Moon's far side or Mars, intelligent personal assistants must be fully functional offline.

**Instructor 3** pointed out that the relevance of an IPA depends strongly on the mission context: *"It largely depends on where your EVA is taking place. If it's on ISS, then maybe not. But if you're on Mars, you'd probably need something like this, because you won't have real-time communication with the mission control centre."*

**Engineer 1** raised technical concerns about embedding IPA systems directly into EVA suits, especially the challenge of implementing reliable AI locally without relying on Earth-based servers: *"But how do you do that (integrating the AI assistant in the suit)? Because then you will need to implement the AI locally, it will need to be either in-suit, or it would have to be on a spacecraft. But ideally in suit, how do you implement something like that in suit?"*

Beyond connectivity, the issue of hardware viability also surfaced. Engineer 1 noted the limitations imposed by energy consumption and computational power, particularly for radiation-hardened environments. Similarly, Astronaut 2 emphasized the importance of local control for mission-critical data and functionality:

**Astronaut 2:** *"I want to have [local] control. I don't want to have to ask the mission control centre, for example, to toggle information in my visor. Because I might not even have a connection to the mission control centre. So all this data, all this functionality, I want to have with me."*

In scenarios such as lunar far-side EVAs or deep cave exploration on Mars, crew members may need to retrieve localized fault diagnosis procedures or verify oxygen reserve levels without ground input. Throughout the interviews, multiple participants pointed out that in such cases, a failure of the IPA could risk the life of the crew. The assistant must thus support autonomous access to important procedural and diagnostic content.

Together, these perspectives underscore that offline operation is not merely a convenience but a mission-critical necessity. They point toward a need for self-contained, resilient, and locally executable IPA systems that can support astronauts autonomously under constrained conditions, without relying on cloud-based processing or external control infrastructure.

## 4.4 Multimodality: Combining Audio and Visual Information

Most participants described voice interaction with an AI assistant as intuitive and hands-free, particularly well suited to the operational realities of space environments. **Scientist 4**, for example, drew inspiration from science fiction portrayals and explained: *"I would prefer to talk... I wouldn't like to have a lot of texts to read when you can just say it."*

However, several participants emphasized that relying solely on audio communication could be limiting, especially during tasks that require situational awareness or procedural clarity. The addition of visual information, either on displays or through AR, was described as important for maintaining efficiency and safety.

**Astronaut 1** noted the risk of reducing information to the audio channel: *"I think that I would not give up augmented reality. You know, voice can be an option to switch features on or off. But it might not always be safe [for more complex tasks] because you completely lose situational awareness if you are just waiting for a voice to respond to you."*

**Engineer 4** highlighted the benefit of visual content for quick reference: *"I'd rather be able to scan what it says."*

**Engineer 3** emphasized the value of visual aids during procedural work: *"You need procedures displayed somewhere... to show you the next three steps."*

**Astronaut 3** envisioned an IPA that could integrate voice commands with augmented reality elements to support lunar surface operations: *"Maybe you have Siri with you and you tell her to put up the map for today's spacewalk, or the road to the next point. Then I would say, 'Hey, can you put up in my head-up display a projection of the track?' And later, 'Okay, now it's in the way, I don't want to see it anymore. You can take it off.' Things like that."*

These perspectives suggest that while audio-based interaction may serve as the primary modality for AI assistants in space, it should be complemented by visual output. This is particularly important for navigation, spatial orientation, and procedural tasks. AR displays were widely seen as a promising solution for presenting such information in a non-intrusive and context-aware manner.

Together, these four themes reflect a shift in how IPAs must be conceptualized within human-systems integration for space: not as convenience tools or as full replacements for human tasks, but as flexible, reliable, multimodal, and offline-capable systems designed to actively support astronaut-centered autonomy and decision-making during missions.

## 5 Discussion

This study examined the requirements and limitations of IPAs for future human spaceflight missions beyond Low Earth Orbit. Based on insights from 15 domain experts, including 3 astronauts, we identified four key requirements: flexibility, reliability, offline availability, and multimodal feedback. These differ meaningfully from terrestrial IPA use cases and are grounded in specific operational contexts such as EVAs, anomaly response, life support monitoring, and navigation.

**Table 2** Comparison of IPA systems across four key criteria derived from expert interviews: offline availability, flexibility, reliability, and multimodal feedback.

| System | Offline Availability | Flexibility | Reliability | Multimodal Feedback |
|---|---|---|---|---|
| CORE | Offline with local LLM, RAG and KG | Dynamic adjustments via vector database, and knowledge graph | Designed for transparency and robustness | Combine audio and AR overlays |
| METIS | Offline with local LLM, RAG and KG | Dynamic adjustments via vector database, and knowledge graph | Designed for transparency and robustness | Text only |
| CIMON | Not offline, cloud dependence | Fixed interaction scripts | Tested in controlled ISS setting, limited robustness | Audio and screen-based interaction |
| Alexa (Artemis I) | Not offline, requires cloud connectivity | Rigid, rule-based structure | Demo only, not reliable for mission use | Audio only |
| AI4U | Some offline capacity with trained modules | Learns tasks during analog mission training | Early stage, no long-term reliability shown | Audio only |

### 5.1 Flexibility in Dynamic Operations

Overall, the study participants expressed a strong preference for IPAs that are highly adaptable in managing tasks whilst catering to a wide range of user needs during future human spaceflight missions. In contrast to terrestrial settings, where assistants support routine activities, future missions will demand on-the-fly adjustments to task sequences, unexpected procedural deviations, and adaptive communication styles.

For example, during an EVA, an astronaut may need to reprioritize steps in a checklist or receive contextual guidance for an unforeseen problem. Current rule-based IPAs are limited in these situations. Therefore, the participants suggested that there is a need for assistants that can understand intent, dynamically restructure procedures, and shift between brief, directive responses and more detailed explanations based on situational or personal needs.

Similar requirements have already been elaborated in other domains. For instance, in an effort to enable Unmanned Aerial Vehicles to better handle unpredictable mission scenarios, Keneni et al. advocated replacing traditional rule-based AI with a hybrid approach that integrates neural networks and fuzzy logic into a single system [20].

In the context of flexibility, prior work on human-systems integration (HSI) has mostly emphasized dynamic function allocation and adaptive control between human and automated systems [15, 35]. These approaches often focus on high-level decisions about when to delegate tasks or shift roles based on cognitive workload or mission state, giving the system a specific degree of flexibility depending on the situation. While valuable, this system-level framing partly differs from the type of flexibility highlighted in our study. As there is a difference between adaptable automation, where the human explicitly controls how automation is applied, and adaptive automation, where the system itself changes its behavior in response to inferred user or task states [11].

In our study, the participants described a need for interaction-level flexibility, namely, IPAs that actively adapts within tasks and requests to user preferences, communication styles, and situational needs, not only to the level of astronaut autonomy. Rather than passively taking over tasks, the IPA must respond to astronaut intent, restructure procedures dynamically, and provide multimodal feedback in a way that enhances and extends, rather than replaces, astronaut autonomy and active decision making, which would represent a form of adaptable automation, enhanced by adaptive elements that respond to inferred intent or context.

This aligns with findings in prior HSI studies, showing that adaptable automation leads to better task performance and reduced workload [11]. Given the safety-critical nature of spaceflight, astronauts may prefer the perceived control offered by adaptable systems, aligning with their desire to direct IPA behavior in context-sensitive ways (e.g., choosing terse vs. verbose feedback, or manually initiating a procedure switch).

**Design Guideline 1.** *IPAs for spaceflight should primarily implement adaptable automation that allows astronauts to manually adjust task flow, modulate interaction style, and influence system behavior in real time. Rather than taking over all decision-making, the assistant should support astronauts by enabling control over task sequences (e.g., skipping, reordering steps) and tailoring feedback (e.g., short vs. detailed explanations, communication style). Rather than replacing astronaut autonomy, adaptable support should be complemented by selective elements of adaptive automation, such as context-sensitive suggestions, provided they remain transparent and user-overridable.*

From a technical standpoint, this could be achieved by integrating a GPT-based language model with a graph retrieval-based framework like GraphRAG [14] as well as steering vectors to for example adjust the language models tone based on the astronauts needs and preferences [22]. Such a system would allow the assistant to retrieve and adjust mission plans on the fly, adapt its interaction style to individual preferences, and support procedural changes without relying on real-time communication with Earth.

## 5.2 Reliability in Safety-Critical Contexts

Reliability was consistently suggested as one of the most important requirements for the use of IPAs in human spaceflight missions. Participants drew attention to past frustrations with unreliable assistants such as CIMON or Alexa, noting delays, misinterpretations, or incorrect outputs. In high-risk operational settings, such failures are not just inconvenient but potentially safety-critical. While users on Earth can fall back on human support when digital assistants fail, astronauts in deep space must rely solely on the system, making the error tolerance effectively zero. As such, participants emphasized that trust is a prerequisite for use. This aligns with prior research, which shows that trust in automation is closely tied to intended usage [12].

Participants also highlighted the importance of explainability, commonly defined as "the ability for the human user to understand the agent's logic" [37, p.678]. Our research suggests, that being able to understand where a recommendation is coming from, how confident the system is, and what reasoning underlies its output could help astronauts to calibrate their trust in the IPA. This, in turn, helps avoid overreliance, while also ensuring that astronauts are not dismissing correct recommendations due to a lack of confidence in the system. These findings are supported by previous work on human-agent teaming and trust calibration [2, 37], and align with NASA's ethical guidelines for AI use, which state that "Solutions must clearly state if, when, and how an AI system is involved, and AI logic and decisions must be explainable" [30, p.3].

**Design Guideline 2.** *In safety-critical space environments, trust and explainability are essential for effective human-agent collaboration. Astronauts must be able to understand and verify the assistant's responses, especially when no human fallback is available. IPAs should therefore provide clear explanations, indicate confidence levels, and make the origin of their recommendations understandable to support explainability and trust calibration.*
From a technical standpoint, this could be achieved by grounding IPA outputs in structured, verifiable knowledge sources, such as KGs and by integrating RAG architectures [14, 26]. These approaches connect the generative capabilities of LLM models with reliable content and support explainability by showing astronauts where the output is coming from.

## 5.3 Offline Availability for Autonomous Operations

Participants stressed that an IPA tailored for space missions would need to function offline, citing the need for local autonomy and reliability as key reasons. The connection between an AI system's accuracy and its perceived trustworthiness is well established [25], and is further amplified in safety-critical contexts [46]. During operations on the Moon's far side or on Mars, communication with Earth will be delayed or unavailable, resulting in a situation where IPAs will have to operate fully offline and deliver reliable functionality without any external humans support (e.g., from mission control centres).

Under such conditions, a system's trustworthiness will inevitably be even more dependent on its perceived reliability. IPAs that fail or degrade under communication constraints could not be relied upon in critical operations. As such, the IPA must be resilient, self-contained, and capable of delivering consistent performance regardless of external connectivity.

**Design Guideline 3.** *In environments where real-time communication with Earth is not possible, such as during Mars missions or lunar operations in shadowed regions, IPAs must function autonomously to support astronaut decision-making and task execution. This requires full offline operability, ensuring that core capabilities remain available even without external support.*
From a technical standpoint, this can be achieved by deploying optimized, hardware-efficient LLMs that utilize reasoning (e.g., Magistral [32]) and combine them with reasoning over knowledge graphs [7]. These approaches allow advanced IPA systems to run locally on wearable or embedded computing platforms, maintaining functionality in communication-constrained conditions.

## 5.4 Multimodal Feedback for Situational Awareness

The participants explained that relying only on audio information could reduce performance, for instance, when astronauts are hands-busy or have limited visual focus, additional output modes become critical. This observation aligns with Multiple Resource Theory [49], which suggests that human attentional capacity is limited and can be supported by distributing information across multiple sensory channels. Accordingly, multimodal interaction has been found to improve spatial orientation [10], safety [44], as well as learning performance of users [40]. AR technology integrated into the astronaut helmet could in this sense allow astronauts to quickly scan instructions, interpret navigation data, or check system diagnostics in a spatial, safe, and intuitive manner, reducing cognitive effort compared to relying on voice or text alone.

Participants described scenarios where audio feedback alone led to confusion or missed instructions. The combination of spoken dialogue with visual overlays could reduce cognitive load and could therefore support better situational awareness. Visually seeing the next three checklist items while hearing confirmations could also create redundancy and therefore reduce the chance of error.

**Design Guideline 4.** *IPAs should integrate with AR systems to provide layered, context-aware information that shows spatial cues, diagnostics, or mission timelines. These should be linked to the underlying knowledge base, allowing astronauts to explore connections or clarify specific items visually. We would furthermore recommend a combination of different modalities, such as voice and sounds, written text, and visual AR output, to enhance the intuitive understanding of data, to reduce cognitive load, and to create redundancy, especially in safety-critical scenarios. Yet, a critical balance should be achieved to only provide the astronaut with necessary and supportive cues to ensure situational awareness and to balance workload.*

## 5.5 Tensions and Trade-offs: Balancing All Four Themes

Our findings, however, do not only show four identified requirements (flexibility, reliability, offline availability, and multimodal information) that should all be fulfilled for space-related IPAs, but also a tension between each of them. Fulfilling one could introduce additional friction for another requirement, requiring a balance between the identified factors.

For instance, participants expressed a strong preference for assistants that can dynamically respond to changing mission conditions or user needs (flexibility). Yet, making the assistant more flexible in its responses could create a higher risk of wrong or unexpected answers affecting its reliability.

Similarly, offline availability was framed as a non-negotiable requirement for missions beyond Low Earth Orbit. However, the need to operate without access to cloud-based resources raises questions about how to support and enhance explainability when the assistant cannot access external databases or justification mechanisms to allow for a high degree of flexibility.

We likewise came across instances of ambivalence surrounding the assistant's interaction style. While the idea of a conversational agent that adapts emotionally to user stress or mood was welcomed in theory, many participants voiced discomfort with systems that attempt to mimic human behavior too closely. This was especially the case in situations involving risk or time pressure, where predictability and control took precedence over social or affective engagement.

Moreover, the integration of visual and auditory modalities was seen as critical to support astronaut performance in cognitively demanding environments. Yet this combination also introduces the risk of sensory overload if not carefully calibrated to task demands. The challenge lies not only in offering multiple channels of information but in ensuring that these are presented in ways that enhance rather than compete for user attention, while ensuring reliability of the system that in an ideal scenario could also react to as many novel situations as possible based on its database while also being available offline.

Therefore, all themes can be seen as important factors that should be in balance with each other to balance competing goals. A one-size-fits-all approach is unlikely to succeed. Instead, systems must be context-aware, adapting their behavior to the astronaut's current needs and situation, workload, and environment while individually adjusting the specific requirement based on its importance in the given situation or mission.

### 5.6 Toward Human-Centered IPA Design for Spaceflight

To summarize, our findings indicate that future IPAs designed for space exploration must explicitly address the requirements of flexibility, reliability, offline availability, and multimodal interaction. The presented insights from astronauts and domain experts emphasize these important design criteria, exceeding the capabilities of traditional rule-based systems and narrow AI. Only the presented approaches (see related work) shown by METIS [6] and CORE [5], integrating KGs [7], generative AI (GPT models) [32], RAG [26], and, for instance AR, could be technical approaches to fulfill the requirements identified in our study (shown in Figure 2). Overall, our study reaffirms the importance of a human-centered approach to space technology development, what we define as a balanced astronaut-oriented design [4]. This approach prioritizes the astronaut's perspective first, while iteratively involving all relevant stakeholders in a co-design process. Rather than retrofitting existing AI systems for extreme environments, IPAs should be co-designed with the end-user in mind, focusing not only on technical feasibility but on the real operational needs, constraints, and preferences of astronauts. Following this approach could also help with balancing trade-offs between the different requirement themes that emerged during the study, for example, because it enables iterative feedback from astronauts during the design process, developers can tailor the assistant's interaction style (flexibility) to different task contexts without compromising trust (reliability), even in offline conditions where fallback options are limited.

## 6 Conclusion and Future Work

Our work underscores the importance of IPA technologies for future space missions, particularly those beyond low Earth orbit, where crews will face complex tasks, high uncertainty, and limited access to ground control. Through semi-structured interviews with astronauts and space experts, we identified key design challenges for conversational agents in this environment and proposed solutions grounded in recent advancements in natural language processing.

Our proposed technical approach centers on a flexible IPA capable of adapting to different tasks and interaction styles, supported by GPT models. Unlike narrow AI systems, these models offer advantages in reasoning, creativity, and handling unforeseen scenarios. Equally important is system reliability, which directly shapes user trust and engagement. To address this, we recommend integrating RAG and knowledge graph technologies, enabling transparent, verifiable outputs during critical operations, such as shown with METIS [6] and CORE [5].

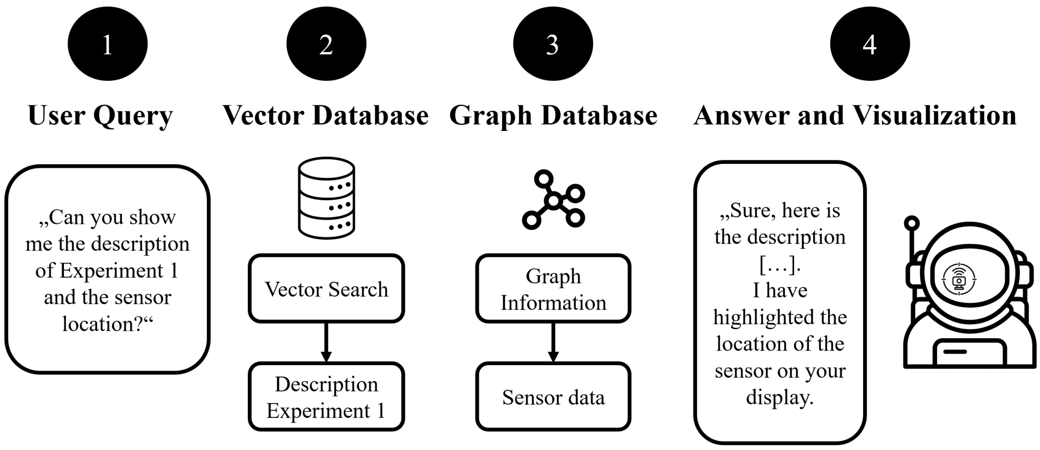


**Figure 2** Example integration of GPT, RAG, KGs and AR.

**Table 3** Design considerations for intelligent personal assistants in spaceflight, organized by core user needs, mission-specific constraints, and actionable recommendations.

| **Theme** | **Representative Use Cases** | **Spaceflight-Specific Challenges** | **Design Recommendations** |
|---|---|---|---|
| Flexibility | EVA procedures, anomaly handling, dynamic checklists | Requires real-time adaptation without ground control input | Support task reprioritization, interaction personalization, and flexible procedure handling through adaptable and selectively adaptive automation |
| Reliability | Life support queries, system diagnostics, checklist execution | Failures may impact safety; no fallback to human assistance in critical moments | Ensure consistent, explainable outputs grounded in trusted data; integrate confidence levels and traceable reasoning to aid trust calibration |
| Offline Availability | Lunar far-side EVAs, Mars surface operations, habitat troubleshooting | Communication blackouts and latency limit support from mission control | IPAs must operate autonomously with locally embedded knowledge, optimized models, and robust decision-making capabilities |
| Multimodal Feedback | Navigation, maintenance workflows, procedural training | Audio-only feedback limits situational awareness during hands-busy tasks | Combine voice with visual cues (e.g., AR overlays, helmet HUDs) to support spatial understanding, reduce cognitive load, and enhance safety |

Offline availability is another essential requirement. Given the communication constraints of lunar and Martian missions, IPA systems must operate autonomously and efficiently without cloud connectivity. Advances in compact, hardware-efficient language models support the feasibility of such offline systems. Finally, to support situational awareness, we propose combining voice-based interaction with visual information presented through AR, allowing astronauts to interpret complex content such as checklists, navigation instructions, or system diagnostics more intuitively.

Looking ahead, our future work focuses on evaluating and refining the CORE system through applied experimentation and user studies. Specifically, we aim to explore the following directions:

- We will integrate the system into a high-fidelity simulation of an Artemis mission. Domain experts will perform EVA scenarios within a virtual reality environment modeled on the Moon's south pole, augmented with AR overlays for navigation and procedural support. We will assess performance across different task contexts and system variants, such as KG plus RAG versus GPT only, or with versus without confidence scores.
- We will investigate optimal strategies for distributing content across visual and auditory channels. This includes identifying which types of information, such as procedures, alerts, or spatial data, are most effectively conveyed through voice or visual displays to minimize cognitive load.
- We will evaluate the robustness of the speech recognition component in acoustically challenging conditions and across diverse speaker profiles. This includes testing and potentially integrating models such as Whisper [36], with a focus on accuracy and responsiveness during simulated EVA activities.
- We will extend the system with computer vision capabilities to support real-time monitoring of both the environment and astronaut activity. This includes hazard detection and geological analysis. Detected features will be visualized through AR overlays to assist decision making in operational settings [39, 16, 3, 47, 29, 41].

## References


1 Amazon. Amazon and Lockheed Martin are sending Alexa to space as part of Callisto, a technology demonstration on NASA's upcoming Artemis I mission. `https://www.aboutamazon.com/news/devices/alexa-take-me-to-the-moon`, 2022. [Online; accessed 22-January-2024].

2 Zahra Atf and Peter R Lewis. Is trust correlated with explainability in ai? a meta-analysis. *IEEE Transactions on Technology and Society*, 2025.

3 Rachel Bellisle, Caroline Bjune, and Dava Newman. Considerations for wearable sensors to monitor physical performance during spaceflight intravehicular activities. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 4160–4164. IEEE, 2020. `doi:10.1109/EMBC44109.2020.9175674`.

4 Leonie Bensch, Tommy Nilsson, Paul de Medeiros, Florian Dufresne, Andreas Gerndt, Flavie Rometsch, Georgia Albuquerque, Frank Ole Flemisch, Oliver Bensch, Michael Preutenborbeck, et al. Towards balanced astronaut-oriented design for future eva space technologies. In *SpaceCHI: 3.0 A Conference on Human-Computer Interaction for Space Exploration*, 2023.

5 Oliver Bensch, Leonie Bensch, Tommy Nilsson, Florian Saling, Bernd Bewer, Sophie Jentzsch, Tobias Hecking, and J. Nathan Kutz. Ai assistants for spaceflight procedures: Combining generative pre-trained transformer and retrieval-augmented generation on knowledge graphs with augmented reality cues, 2024. `doi:10.48550/arXiv.2409.14206`.

6 Oliver Bensch, Leonie Bensch, Tommy Nilsson, Florian Saling, Wafa M. Sadri, Carsten Hartmann, Tobias Hecking, and J. Nathan Kutz. Towards a reliable offline personal ai assistant for long duration spaceflight, 2024. doi:10.48550/arXiv.2410.16397.

7 Oliver Bensch and Tobias Hecking. Towards open domain literature based discovery. In *3rd International Open Search Symposium*, 2021.

8 Adam M. Braly, Benjamin Nuernberger, and So Young Kim. Augmented reality improves procedural work on an international space station science instrument. *Human factors*, 61(6):866–878, 2019. Publisher: SAGE Publications Sage CA: Los Angeles, CA. doi:10.1177/0018720818824464.

9 Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

10 Verena N Buchholz, Samanthi C Goonetilleke, W Pieter Medendorp, and Brian D Corneil. Greater benefits of multisensory integration during complex sensorimotor transformations. *Journal of neurophysiology*, 107(11):3135–3143, 2012.

11 Gloria Calhoun. Adaptable (not adaptive) automation: Forefront of human–automation teaming. *Human factors*, 64(2):269–277, 2022. doi:10.1177/00187208211037457.

12 Hyesun Choung, Prabu David, and Arun Ross. Trust in ai and its role in the acceptance of ai technologies. *International Journal of Human–Computer Interaction*, 39(9):1727–1739, 2023. doi:10.1080/10447318.2022.2050543.

13 Deutsches Zentrum für Luft- und Raumfahrt (DLR). CIMON gelingt Weltpremiere mit Alexander Gerst auf der ISS. https://www.dlr.de/de/aktuelles/nachrichten/2018/4/20181116_cimon-weltpremiere-alexander-gerst-iss, November 2018. Accessed: 2024-01-24.

14 Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, Dasha Metropolitansky, Robert Osazuwa Ness, and Jonathan Larson. From local to global: A graph rag approach to query-focused summarization, 2025. arXiv:2404.16130.

15 MA Goodrich and ML Cummings. A human factors perspective on next generation unmanned aerial systems. *Int C2 J*, 1(2):1–24, 2007.

16 Brian HW Guo, Yang Zou, Yihai Fang, Yang Miang Goh, and Patrick XW Zou. Computer vision technologies for safety science and management in construction: A critical review and future research directions. *Safety science*, 135:105130, 2021.

17 NC Haney and TG Graff. Apollo 17 eva-1 and eva-2 task decomposition: Planning for artemis and future mars missions. In *Lunar and Planetary Science Conference*, 2020.

18 Matthew H Hersch. Checklist: The secret life of apollo's 'fourth crewmember'. *The Sociological Review*, 57(1_suppl):6–24, 2009.

19 Gerda Horneck, Rainer Facius, Michael Reichert, Petra Rettberg, Wolfgang Seboldt, Dietrich Manzey, Bernard Comet, A Maillet, H Preiss, L Schauer, et al. Humex, a study on the survivability and adaptation of humans to long-duration exploratory missions, part i: lunar missions. *Advances in Space Research*, 31(11):2389–2401, 2003.

20 Blen M Keneni, Devinder Kaur, Ali Al Bataineh, Vijaya K Devabhaktuni, Ahmad Y Javaid, Jack D Zaientz, and Robert P Marinier. Evolving rule-based explainable artificial intelligence for unmanned aerial vehicles. *IEEE Access*, 7:17001–17016, 2019. doi:10.1109/ACCESS.2019.2893141.

21 Natalie M Kintz and Lawrence A Palinkas. Communication delays impact behavior and performance aboard the international space station. *Aerospace Medicine and Human Performance*, 87(11):940–946, 2016.

22 Kai Konen, Sophie Jentzsch, Diaoulé Diallo, Peer Schütt, Oliver Bensch, Roxanne El Baff, Dominik Opitz, and Tobias Hecking. Style vectors for steering generative large language model, 2024. doi:10.48550/arXiv.2402.01618.

23 Markus Landgraf, Jennifer Reynolds, Naoki Sato, Kandyce E. Goodliff, Clark Esty, and Martin Picard. Lunar Surface Concept of Operations for the Global Exploration Roadmap Lunar Surface Exploration Scenario, October 2021. NTRS Author Affiliations: European Space Agency, Japan Aerospace Exploration Agency, Langley Research Center, Johnson

Space Center, Canadian Space Agency NTRS Meeting Information: 72nd International Astronautical Congress 2021; 2021-10-25 to 2021-10-29; undefined NTRS Document ID: 20210022596 NTRS Research Center: Headquarters (HQ). URL: `https://ntrs.nasa.gov/citations/20210022596`.

**24** Lauren Blackwell Landon, Kelley J Slack, and Jamie D Barrett. Teamwork and collaboration in long-duration space missions: Going to extremes. *American Psychologist*, 73(4):563, 2018.

**25** Markus Langer, Cornelius J König, Caroline Back, and Victoria Hemsing. Trust in artificial intelligence: Comparing trust processes between human and automated trustees in light of unfair bias. *Journal of Business and Psychology*, 38(3):493–508, 2023.

**26** Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks, 2021. `arXiv:2005.11401`.

**27** Paul Pu Liang, Karan Ahuja, and Yiyue Luo. Multimodal ai for human sensing and interaction. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–4, 2025. `doi:10.1145/3706599.3706651`.

**28** Thomas Marshburn, Mihriban Whitmore, Rosie Ortiz, Michele Segal, Kieran Smart, and Catherine Hughes. An independent human factors analysis and evaluation of the emergency medical protocol checklist for the international space station. Technical report, NASA, 2003.

**29** Patrick Charles McGuire, Enrique Díaz-Martínez, Jens Ormö, Javier Gómez-Elvira, José Antonio Rodriguez-Manfredi, Eduardo Sebastián-Martínez, Helge Ritter, Robert Haschke, Markus Oesker, and Jörg Ontrup. The cyborg astrobiologist: scouting red beds for uncommon features with geological significance. *International Journal of Astrobiology*, 4(2):101–113, 2005.

**30** Edward McLarney, Yuri Gawdiak, Nikunj Oza, Chris Mattmann, Martin Garcia, Manil Maskey, Scott Tashakkor, David Meza, John Sprague, Phyllis Hestnes, et al. Nasa framework for the ethical use of artificial intelligence (ai). *NASA Report*, 2021.

**31** David A Mindell. *Digital Apollo: Human and machine in spaceflight.* IEEE Annals of the History of Computing, 2011.

**32** Mistral-AI, :, Abhinav Rastogi, Albert Q. Jiang, Andy Lo, Gabrielle Berrada, Guillaume Lample, Jason Rute, Joep Barmentlo, Karmesh Yadav, Kartik Khandelwal, Khyathi Raghavi Chandu, Léonard Blier, Lucile Saulnier, Matthieu Dinot, Maxime Darrin, Neha Gupta, Roman Soletskyi, Sagar Vaze, Teven Le Scao, Yihan Wang, Adam Yang, Alexander H. Liu, Alexandre Sablayrolles, Amélie Héliou, Amélie Martin, Andy Ehrenberg, Anmol Agarwal, Antoine Roux, Arthur Darcet, Arthur Mensch, Baptiste Bout, Baptiste Rozière, Baudouin De Monicault, Chris Bamford, Christian Wallenwein, Christophe Renaudin, Clémence Lanfranchi, Darius Dabert, Devon Mizelle, Diego de las Casas, Elliot Chane-Sane, Emilien Fugier, Emma Bou Hanna, Gauthier Delerce, Gauthier Guinet, Georgii Novikov, Guillaume Martin, Himanshu Jaju, Jan Ludziejewski, Jean-Hadrien Chabran, Jean-Malo Delignon, Joachim Studnia, Jonas Amar, Josselin Somerville Roberts, Julien Denize, Karan Saxena, Kush Jain, Lingxiao Zhao, Louis Martin, Luyu Gao, Lélio Renard Lavaud, Marie Pellat, Mathilde Guillaumin, Mathis Felardos, Maximilian Augustin, Mickaël Seznec, Nikhil Raghuraman, Olivier Duchenne, Patricia Wang, Patrick von Platen, Patryk Saffer, Paul Jacob, Paul Wambergue, Paula Kurylowicz, Pavankumar Reddy Muddireddy, Philomène Chagniot, Pierre Stock, Pravesh Agrawal, Romain Sauvestre, Rémi Delacourt, Sanchit Gandhi, Sandeep Subramanian, Shashwat Dalal, Siddharth Gandhi, Soham Ghosh, Srijan Mishra, Sumukh Aithal, Szymon Antoniak, Thibault Schueller, Thibaut Lavril, Thomas Robert, Thomas Wang, Timothée Lacroix, Valeriia Nemychnikova, Victor Paltz, Virgile Richard, Wen-Ding Li, William Marshall, Xuanyu Zhang, and Yunhao Tang. Magistral, 2025. `arXiv:2506.10910`.

**33** Nasa. Apollo 11 Preliminary Science Report. `https://history.nasa.gov/alsj/a11/a11psr.html`, 1969. Accessed: 2023-11-18.

**34** Gregory Navarro, Marie-Christine Desjean, and Alexis Paillet. Eclss technology roadmap at spaceship fr. In *N/A*. 2023 International Conference on Environmental Systems, 2023.

35 Raja Parasuraman, Keryl A Cosenzo, and Ewart De Visser. Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology*, 21(2):270–297, 2009.

36 Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

37 Avi Rosenfeld and Ariella Richardson. Explainability in human–agent systems. *Autonomous agents and multi-agent systems*, 33:673–705, 2019. doi:10.1007/S10458-019-09408-Y.

38 Hans-Christian Schmitz, Frank Kurth, Kevin Wilkinghoff, Uwe Mullerschkowski, Christian Karrasch, and Volker Schmid. Towards robust speech interfaces for the iss. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion*, pages 110–111, 2020. doi:10.1145/3379336.3381496.

39 JoonOh Seo, SangUk Han, SangHyun Lee, and Hyoungkwan Kim. Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, 29(2):239–251, 2015. doi:10.1016/J.AEI.2015.02.001.

40 Ladan Shams and Aaron R Seitz. Benefits of multisensory learning. *Trends in cognitive sciences*, 12(11):411–417, 2008.

41 Helia Sharif, Maxim Ralchenko, Claire Samson, and Alex Ellery. Autonomous rock classification using bayesian image analysis for rover-based planetary exploration. *Computers & Geosciences*, 83:153–167, 2015. doi:10.1016/J.CAGEO.2015.05.011.

42 Elizaveta Shashkovaa, Gregory Navarrob, Raphaëlle N Royc, Alexis Pailletd, and Luc Truntzlere. Study and development of an ai assistant for future moon and mars stations. *International Astronautical Congress*, 2022.

43 Marshall Smith, Douglas Craig, Nicole Herrmann, Erin Mahoney, Jonathan Krezel, Nate McIntyre, and Kandyce Goodliff. The artemis program: An overview of nasa's activities to return humans to the moon. In *2020 IEEE Aerospace Conference*, pages 1–10. IEEE, 2020.

44 Charles Spence and Cristy Ho. Multisensory interface design for drivers: past, present and future. *Ergonomics*, 51(1):65–70, 2008.

45 SPooN. AI-4U: A personal assistant for astronauts. https://www.spoon.ai/cnes-ai-4u, 2024. Accessed: 2024-01-22.

46 Chris Thames and Yifan Sun. A survey of artificial intelligence approaches to safety and mission-critical systems. In *2024 Integrated Communications, Navigation and Surveillance Conference (ICNS)*, pages 1–12. IEEE, 2024.

47 Leonardo Turchi, Samuel J Payler, Francesco Sauro, Riccardo Pozzobon, Matteo Massironi, and Loredana Bessone. The electronic fieldbook: A system for supporting distributed field science operations during astronaut training and human planetary exploration. *Planetary and Space Science*, 197:105164, 2021.

48 RC Weber, SJ Lawrence, BA Cohen, JE Bleacher, JW Boyce, MR Collier, D Draper, AL Fagan, CI Fassett, L Gaddis, et al. The artemis iii science definition team report. In *52nd Lunar and Planetary Science Conference*, page 1261, 2021.

49 Christopher D Wickens. Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2):159–177, 2002.