Human-AI Interaction in Space: Insights from a Mars Analog Mission with the Harmony Large Language Model

Hippolyte Hilgers
Université catholique de Louvain, MARS-UClouvain Space Mission, B-1348 Louvain-la-Neuve, Belgium

Jean Vanderdonckt[1]
Université catholique de Louvain, Louvain Research Institute in Management and Organizations (LouRIM) and Institute for Information and Communication Technologies, Electronics and Applied Mathematics (ICTeam), Place des Doyens, 1, B-1348 Louvain-la-Neuve, Belgium

Radu-Daniel Vatavu
MintViz Lab, MANSiD Research Center, Ştefan cel Mare University of Suceava, 13 Universitatii, 720229 Suceava, Romania

Abstract

The operational complexities of space missions require reliable, context-aware technical assistance for astronauts, especially when technical expertise is not available onboard and communication with Earth is delayed or limited. In this context, Large Language Models present a promising opportunity to augment human capabilities. To this end, we present Harmony, a model designed to provide astronauts with real-time technical assistance, fostering human-AI collaboration during analog missions. We report empirical results from an experiment involving seven analog astronauts that evaluated their user experience with Harmony in both a conventional environment and an isolated, confined, and extreme physical setting at the Mars Desert Research Station over four sessions, and discuss how the Mars analog environment impacted their experience. Our findings reveal the extent to which human-AI interactions evolve across various user experience dimensions and suggest how Harmony can be further adapted to suit extreme environments, with a focus on SpaceCHI.

2012 ACM Subject Classification Human-centered computing → Field studies, User studies, HCI design and evaluation methods, Interaction paradigms, Graphical user interfaces; Applied computing → Aerospace; Computing methodologies → Machine learning approaches; Information systems → Language models; Applied computing → Computer-assisted instruction

Keywords and phrases Extreme user experience, Human-AI interaction, Isolated-confined-extreme environment, Interaction design, Large Language Models, Mars Desert Research Station, Space mission, Technical assistance, Technical documentation, User experience

Digital Object Identifier 10.4230/OASIcs.SpaceCHI.2025.1

Supplementary Material www.kaggle.com/datasets/jeanvanderdonckt/dataset-spacechi2025-paper

Funding *Hippolyte Hilgers*: MARS-UCLouvain, Atlas Crew 2024.
Jean Vanderdonckt: Novel Extended Reality Models and Software Framework for Interactive Environments with Extreme Conditions (XXR) project, funded by Wallonnie-Bruxelles Interactional (WBI) under Grant no. 650763.
Radu-Daniel Vatavu: Novel Extended Reality Models and Software Framework for Interactive Environments with Extreme Conditions (XXR), Romanian Ministry of Research, Innovation and Digitization, CCCDI-UEFISCDI, PN-IV-P8-8.3-PM-RO-BE-2024-0003, within PNCDI IV.

[1] Corresponding author

Figure 1 Photograph taken during our analog mission conducted at the Mars Desert Research Station, depicting an ICE environment with challenging physical constraints.

1 Introduction

Isolated, Confined, and Extreme (ICE) environments [2, 23] refer to settings with challenging constraints for their inhabitants, including psychological [27, 54], physical [29, 61], social [41], cognitive [28], and technological [3, 20], such as in space [23] or research stations [28]. In these environments, users are often isolated [2] or in small teams [25], typically confined in enclosed spaces for prolonged periods [26], and constrained by protective equipment.

These extreme conditions pose new challenges to the user experience (UX) of interacting with computer systems [54]. Although extensive research in Human-Computer Interaction (HCI) has covered traditional contexts of use [10, 13, 57], limited work has addressed UX challenges in ICE environments, where the extreme nature of the environment and the remote habitat [23] can significantly impact UX [52]. Under these constraints, astronauts must perform a wide range of tasks [4], from operating scientific instruments [60] to managing life-support systems [30], which require access to readily-available technical knowledge and documentation [14], not always present onboard. Moreover, astronauts cannot always count on real-time technical support from Earth given communications latency [5]. As Large Language Models (LLMs) can provide context-aware, real-time support for many tasks, such as self-scheduling [52] and operational exploration [21], they represent promising candidates for deployment as space mission assistants [5, 6]. This role makes them particularly relevant for ICE environments, where human-AI interaction remains underexplored.

In this paper, we examine how a mission at the Mars Desert Research Station, a representative ICE environment (Figure 1), affects the UX of analog astronauts interacting with Harmony, an LLM developed to provide technical assistance during space missions.

2 Related Work

Previous research about ICE environments has reported on individual attributes for contextual adaptation, including emotional stability, self-control, and task-oriented coping [2], while environmental design has led to recommendations involving display technology [17], personalization, and areas fostering privacy and socialization [54]. Moreover, environmental factors, such as sensory deprivation [61], sleep disturbances [63], and group dynamics, were shown to impact the functioning of ICE sojourners [56]. In this context, we relate to scientific literature at the intersection of HCI and space exploration with a focus on human-AI interaction.

2.1 HCI in Space

SpaceCHI [44, 45, 62] represents an initiative of the HCI community to investigate interactive computer systems in space, as a representative ICE environment, by "designing new types of interactive systems and computer interfaces that can support human living and working in space and elsewhere in the solar system" [44, p. 1]. SpaceCHI emphasizes the diversity of topical coverage in space exploration requiring HCI knowledge and expertise [3], including software for crew collaboration [53], mission planning [62], human-system resilience and design for maintainability in space [36], participatory design for space systems engineering [39], food experience design for space travel [40], and examinations of the influence of extraterrestrial conditions on designing interactive systems [16, 17, 29].

However, the UX of interactions with computer systems in space has been briefly addressed in the scientific literature. For example, a self-scheduling application evaluated during a Mars analog mission [48] reported UX varying across pragmatic and hedonic dimensions, while Graphical User Interfaces (GUIs) for astronauts have started to be the subject of systematic UX design and assessment [19, 52]. Other research has looked at specific interactive computer technology in space. For example, interviews with astronauts and space experts regarding the capabilities of virtual environments facilitates user-centered approaches to operational performance [39]; MoonBuddy [7] consists of a voice-based VR system for extravehicular activities (EVA); Telemetron [16] is a musical instrument to play in ICE environments with low gravity; and Minerva [11] applies user-centered design for human exploration.

2.2 Human-AI Interaction in Space

AI technologies are increasingly utilized in space stations to enhance crew safety, productivity, and autonomy [15, 43]. Recent research has explored Human-AI interaction (HAI) in various contexts of use, including space applications. For example, a semi-formal representation of HAI using a set of interaction primitives and patterns was proposed in [55] with applications to text summarization [8] and, at the International Space Station (ISS), AI models have proven valuable for data analysis and automation [42]. Specific AI technologies deployed at the ISS include NASA's Robonauts for assisting crew members and ATLAS for asteroid detection. AI has also been explored for providing medical aid to astronauts, such as in the absence of a doctor for the treatment of a fractured tibia [31], data analysis [42], and support for exploratory operations [21]. In this context, critical HAI aspects emphasize the need for mutual trust and reliability [43]. Since astronauts cannot access and manage the vast and diverse knowledge required to conduct space missions, AI assistants [5, 8, 12, 15, 21, 42, 43, 55] could help with finding answers to their questions in context. For example, Bensch et al. [5] combined information retrieval techniques, knowledge graphs, and Augmented Reality (AR)

cues in their AI assistant for spaceflight operations [6]. These advancements enhance astronaut autonomy [4] while simultaneously addressing safety and assurance considerations [43], critical for practical implementations.

2.3 Summary

The recent interest of the HCI community in contributing to humanity's quest for reaching [33], working [30], and living [56] in ICE environments for space exploration [62], sets the context for our examination reported in this paper, where we focus on the UX of analog astronauts interacting with AI assistants in such environments. While proper UX design [22] can optimize the interaction between users and systems in conventional contexts, it requires careful and extensive examination in extraterrestrial environments [38]. Related research has primarily focused on technical aspects, such as hardware and software reliability or communications latency [44, 45, 62, 9] and, when tackling UX, on effectiveness and performance aspects. In this context, the need for research on how the specific challenges of ICE environments impact interactions and UX remains largely unaddressed [59]. Moreover, the growing interest in integrating AI in applications demands examination of the UX resulting from increasingly frequent human-AI interactions. Our work lies at the intersection of SpaceCHI, AI, and UX.

3 Implementation of Harmony

Harmony is an LLM designed to provide astronauts with technical assistance regarding tasks for which they do not necessarily have expertise or for which expertise is not readily available via remote communication with Earth. These tasks may include accessing system documentation, troubleshooting (diagnosing and resolving technical issues with spacecraft systems), experiment guidance (providing step-by-step instructions for scientific experiments), health monitoring (supporting medical procedures and monitoring crew well-being), mission planning (assisting in dynamic re-scheduling and optimizing resource utilization), and self-scheduling (self management of tasks); see details in [48, 52].

3.1 Software Architecture

The software architecture of Harmony is composed of four modules, each responsible for a specific task in the processing pipeline (Figure 2 and Table 1), as follows:

1. **Audio Capture**: Audio input is captured using a microphone and activated with a mouse click acting as a push-to-talk control. This interaction simplifies hardware requirements and allows for flexible input in low gravity conditions [17].
2. **Speech-to-Text Transcription**: The OpenAI Whisper ASR model, a robust and accurate English speech recognition engine, runs locally and transcribes the astronaut's question into text without requiring an Internet connection.
3. **Query Generation and Answering**: The transcribed question is processed and sent to an LLM based on GPT-3.5-turbo, which formulates a concise, mission-adapted response. This module operates either locally, with a fine-tuned model, or via the Google Cloud API when a communication link with Earth is available.[2]

[2] An LLM capable of running locally on the onboard computer was initially implemented but, during the mission, we considered an LLM based directly on GPT-3.5-turbo for faster and more relevant responses.

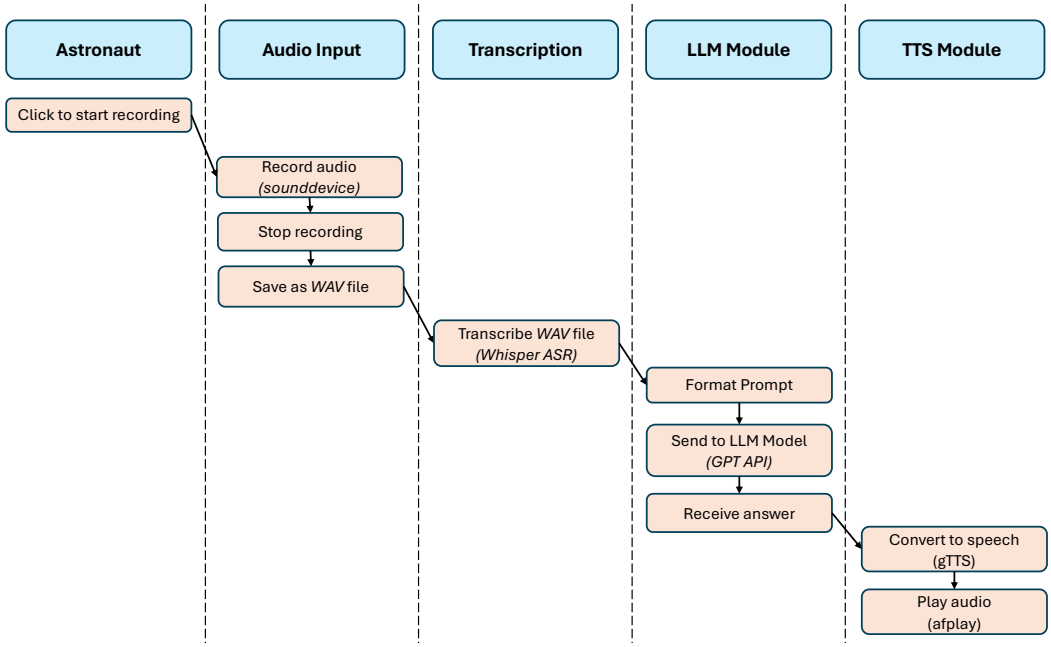

Figure 2 The UML Activity Diagram of Harmony.

Table 1 Overview of the software modules implemented for voice-based interaction in Harmony.

Module	Library/Tool	Function
Audio Recording	`sounddevice`, `pynput`	Capture of voice input via click trigger
Speech-to-Text	`Whisper (base)`	Transcription of recorded audio into text
LLM Interaction	`OpenAI API (GPT-3.5-turbo)`	Generating responses based on transcription
Text-to-Speech	`gTTS (Google Text-to-Speech)`	Conversion of LLM response into vocal output
Audio Playback	`os.system + afplay`	Local playback of the generated audio

4. **Text-to-Speech Response**: The generated answer is converted back into audio using the Google Text-to-Speech gTTS engine, a Python library and Command-Line Interface tool to interface with gTTS API, allowing astronauts to receive direct vocal feedback.

This modular architecture ensures that each component can be independently updated or adapted based on mission requirements. For example, the LLM backend could be swapped for an onboard model in deep space scenarios with communication latency, while a more advanced Text-To-Speech module, such as Tacotron 2 or Coqui TTS, added to the architecture to improve perceived naturalness of voice-based feedback.

3.2 System Walkthrough

To foster trust and reliability [43], Harmony employs a conversational user interface [47] designed based on the following principles [1]: *transparency* to explain the reasoning behind the provided suggestions to ensure astronaut confidence, *adaptability* to personalize responses based on individual crew preferences and roles, and *error management* to detect and mitigate inaccuracies through validation mechanisms, e.g., reinforcing or deforcing an answer. Figure 3 presents a concrete system walkthrough illustrating how Harmony operates in a real-world scenario. This step-by-step description follows the typical interaction between astronauts and Harmony, from voice input to a vocal response: a context of is first provided to the

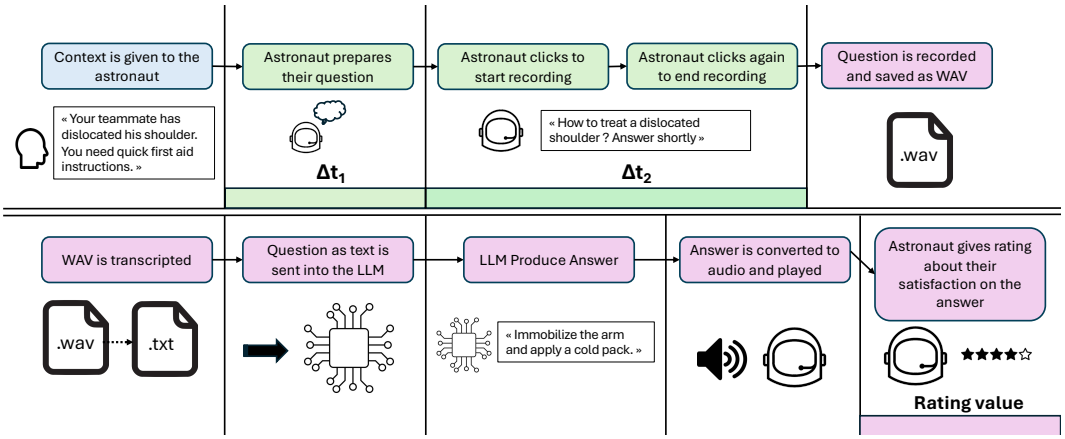

Figure 3 A system walkthrough of Harmony; see Figure 4 for an example.

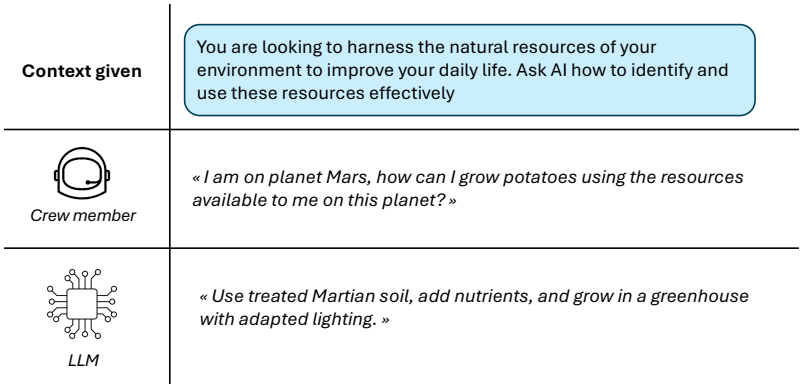

Figure 4 Example of an interaction with Harmony.

analog astronaut with a goal to achieve; subsequently, the astronaut formulates mentally a question corresponding to the goal (Δt_1); the astronaut enters the question with voice input (Δt_2); the question is transcripted and sent to the LLM to produce an answer that is converted to audio to preserve the conversational style; lastly, the astronaut rates their satisfaction regarding the provided answer to improve the LLM training.

4 Experiment

This section presents the experiment conducted to assess the UX of Harmony; see Figure 5 .

4.1 Location

The experiment took place at the Mars Desert Research Station (MDRS), a simulated Mars-inhabited environment in Hanksville, UT, USA, serving as an international research facility for studying human factors and conducting experiments for future Mars missions [18, 46]. MDRS is composed of several modules, including the habitat, the EVA preparation room, and the science dome with a science laboratory; see Figure 5 and a 3D navigation in VR.

4.2 Participants

The experiment was carried out with a crew of seven analog astronauts (four women and three men), aged between 21 and 31 years (M=24.75, SD=2.63), who spent two weeks at the

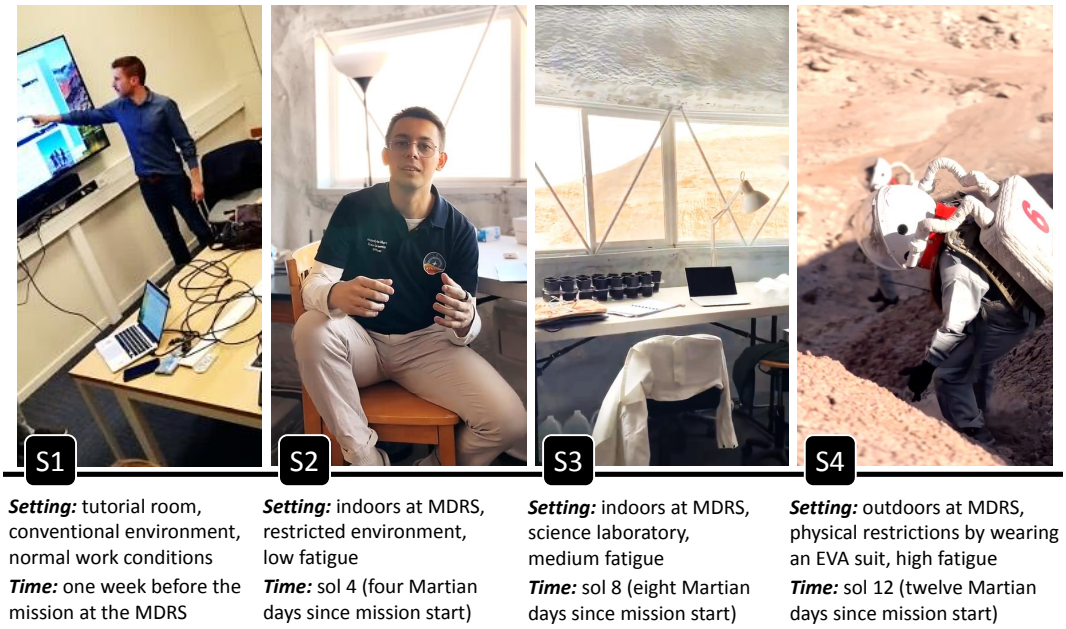

Figure 5 Timeline of the experiment with four stages taking place before (S1) and during the analog mission (S2 to S4) at the MDRS.

MDRS with multiple daily reporting and protocols. Participants had different backgrounds, such as astronomy, biology, chemistry, computer science, geology, and engineering.

4.3 Stimuli

Based on [48, 52], we defined a series of contexts of use, each associated with a situation and a corresponding question, as follows (see Figure 4 for an example):

- Q_1. *Medical assistance in case of injury*: Your teammate has just dislocated their shoulder on a mission. Formulate a question to the AI to learn how to react and what first aid measures to apply.
- Q_2. *Cooking in space with dehydrated products*: You want to prepare pancakes using only dehydrated ingredients. Ask the AI for a suitable recipe.
- Q_3. *Entertainment*: You want the AI to play the role of a general knowledge quiz host. Formulate a question so that it suggests an interesting challenge.
- Q_4. *Communication with a foreign colleague*: You meet a foreign colleague and have difficulty communicating with them. Ask the AI for translation or communication tips.
- Q_5. *Recipe ideas with limited resources*: You have harvested basil and you want to use it to prepare your meal. Ask the AI for two or three dishes ideas.
- Q_6. *Language learning in your spare time*: You want to learn a few words in a new language in your spare time. Ask the AI a question, which will teach you useful vocabulary.
- Q_7. *Physical exercise in a confined space*: You are confined to a small space for a long time and want to keep fit. Ask the AI for a physical exercise routine that can be performed without specialized equipment.
- Q_8. *Project management skills development*: You want to improve your project management skills in your spare time. Ask the AI for resources or a learning plan.

- Q_9. *Using natural resources*: You are looking to harness the natural resources of your environment. Ask the AI how to identify and use these resources effectively.
- Q_{10}. *Team building*: To strengthen team cohesion, you want to organize a team-building game. Ask the AI a question to get ideas tailored to your mission.
- Q_{11}. *Water purification in the event of a technical problem*: A technical problem affects your water supply. Ask the AI what alternative purification methods you can implement.
- Q_{12}. *Night navigation without instruments*: You are lost without a compass and want to find your way using the constellations. Ask the AI a question to learn the basics of astronomical navigation.
- Q_{13}. *Setting up a recycling or composting system*: With limited resources, you want to set up a small recycling or composting system. Ask the AI for advice on how to do it.
- Q_{14}. *Stress management and relaxation*: You are experiencing a lot of stress and would like to practice a relaxation technique. Ask the AI for instructions for a relaxation session.
- Q_{15}. *Recognizing and treating hypothermia*: You are confronted with extremely cold conditions and suspect hypothermia. Ask the AI how to identify symptoms and what treatment to apply with the available means.

4.4 Measures

To evaluate UX, we relied on the UEQ+ method, a modular extension of the User Experience Questionnaire (UEQ) [50, 51], which covers both pragmatic and hedonic UX dimensions and is supported by analysis instruments and published norms [37] for interpreting results [24]. Based on prior work [58, 59], we selected the following UX dimensions for our experiment:

- *Attractiveness*: The overall impression concerning Harmony. Do users like it or not?
- *Efficiency*: The impression that tasks can be successfully performed with Harmony without unnecessary effort. Can users solve their tasks efficiently?
- *Perspicuity*: The impression that Harmony is easy to learn how to use. To what extent do users find Harmony easy to learn?
- *Dependability*: The impression to be in control of the interaction with Harmony. To what extent does Harmony give users the feeling that they are in control of the interaction?
- *Stimulation*: The impression that Harmony is interesting and fun to use. Do users find Harmony exciting and motivating?
- *Novelty*: The impression that Harmony is creative and original. To what extent do users appreciate Harmony as creative? Does it catch their interest?
- *Adaptability*: The impression that Harmony can be easily adapted to personal preferences or working styles. To what extent does Harmony appear adaptable?
- *Trust*: The impression of the users that their data are safe with Harmony and not misused to harm them. Do users feel confident that their data is secure and handled appropriately?
- *Usefulness*: The impression that using Harmony is beneficial. Do users perceive any advantages in interacting with Harmony?
- *Value*: The impression that Harmony looks professional and valuable. To what extent does the design of Harmony convey a sense of professionalism and quality?
- *Visual aesthetics*: The impression that the graphical interface of Harmony looks beautiful and appealing. Do users find the visual design of Harmony attractive and engaging?
- *Intuitive use*: The impression that Harmony can be used immediately without any training or help. Can users start using it right away without needing help or instruction?
- *Trustworthiness of content*: The impression that the information provided by Harmony is of good quality and reliable. To what extent is the information provided by Harmony perceived as accurate and trustworthy?

We refer to UEQ+ [50, 51, 37, 24] for further details about these dimensions. Following prior research [59, 58], we report for each dimension the Scale-Mean-Score (SMS) as the average score obtained on all its subscales, ranging from -3 (negative experience) to $+3$ (positive experience), and the Scale-Mean-Importance (SMI), representing the average weight of importance of a given UX dimension; see Vanderdonckt et al. [58] for calculation details. Furthermore, to compare the scale of a target, *e.g.*, an extreme condition such as Mars, to the corresponding scale of a baseline such as Earth, we use:

$$\text{Scale-Mean-Ratio} = \frac{\text{SMS } (target)}{\text{SMS } (baseline)} \tag{1}$$

$$\text{Scale-Importance-Ratio} = \frac{\text{SMI } (target)}{\text{SMI } (baseline)} \tag{2}$$

Participants' answers were computed with the UEQ data analysis tool and interpreted according to Schrepp et al.'s [51] recommendation: "the standard interpretation of the scale means that values between -0.8 and 0.8 represent a neutral evaluation of the corresponding scale, values superior to 0.8 represent a positive evaluation."

We also measured:

- Quality-Score: a numerical variable representing the perceived quality of the answer provided by Harmony to the question prompted, from 0 (failure) to 1 (lowest quality) to 5 (highest quality for complete and correct answers).
- Success-Factor: a numerical variable defined as the number of iterations needed to complete a task successfully.
- Thinking-Time: a numerical variable defined as the time elapsed between the moment when the task was presented and the moment when the participant started the interaction with Harmony, measured in seconds with a stopwatch (Δt_1 in Figure 3).
- Production-Time: a numerical variable defined as the time needed to formulate and record the question, measured in seconds with a stopwatch (Δt_2 in Figure 3).
- Confidence: an integer variable expressing the degree of confidence attributed to the answer provided by Harmony for the given context, ranging from 1% (no confidence) to 100% (maximum confidence).

4.5 Apparatus and Tasks

Harmony ran on an Apple MacBook Air with a 13" Retina screen (2560×1600 pixels), 8-core 3.2 GHz CPU, 8 GB RAM, and 256 GB SSD. The Apple AirPods Pro 2 were used for audio input and output. We devised four tasks to be carried out in the experiment:

- A *discovery task*, in which the participants received a tutorial on using Harmony, lasting between 5 and 15 minutes, and then interacted freely for another 10 to 15 minutes.
- A *practice task*, in which the participants were instructed to perform a representative task with Harmony to become familiar with it, lasting about 20 minutes.
- A *domain task*, in which the participants received five stimuli (Q_1 to Q_5, Q_6 to Q_{10}, and Q_{11} to Q_{15} in random order). We suggested, but not imposed, a time limit of 10 minutes.
- An *evaluation task*, in which the participants filled out the UEQ+ questionnaire.

The discovery and practice tasks were performed once before the mission in a dedicated tutorial room (see Figure 5, left). The domain and evaluation tasks were performed four times during four subsequent sessions, as follows: a first session S1 one week before the

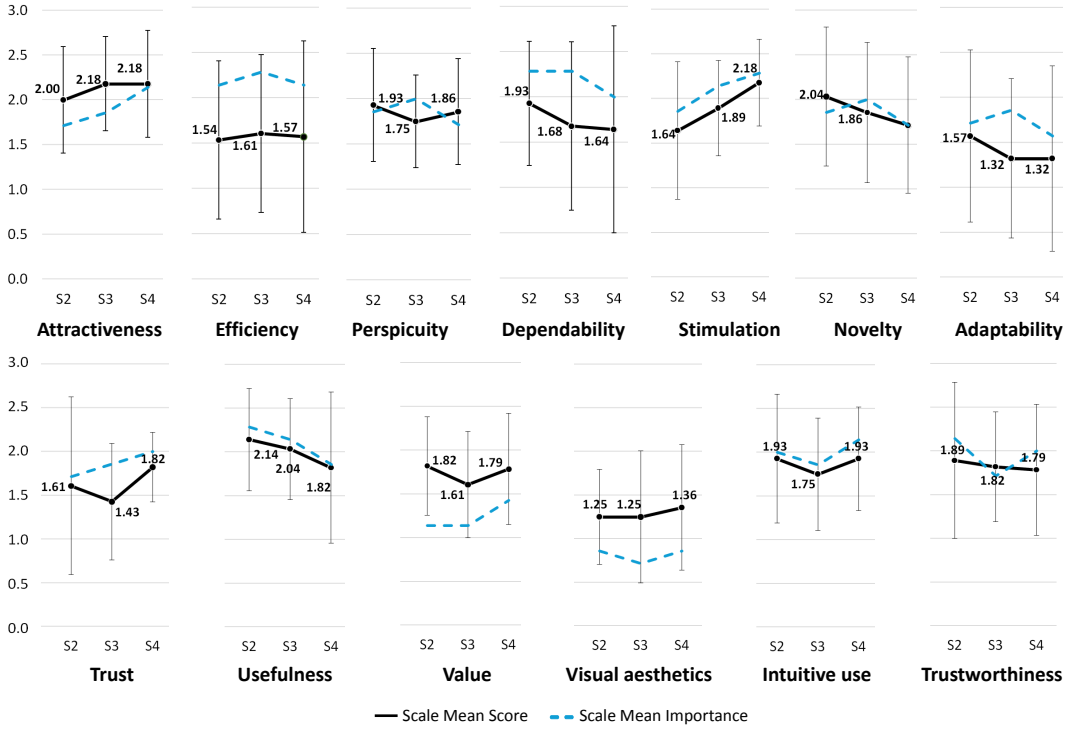

Figure 6 The UX dimensions evaluated for Harmony, showing mean scale scores (solid lines) and scale importance scores (dotted lines) across the mission sessions. The error bars show 95% CIs.

mission in the tutorial room (Figure 5, left), a second session S2 after four Mars days[3](Sol 4) in the science laboratory (Figure 5, middle left), a third session S3 after eight Mars days (Sol 8) outside the station and involving light equipment (Figure 5, middle right), and a fourth session S4 after twelve Mars days (Sol 12) outside the station involving heavy equipment, stress, and high fatigue due to the ICE environment and continuous exposure to its constraints (Figure 5, right).

5 Results and Discussion

5.1 The User Experience of Interacting with Harmony

The panel charts in Figure 6 present the Scale-Mean-Scores (SMS) and Scale-Mean-Importance (SMI) for the various UX dimensions. Overall, all scores fall into the positive experience zone, above the 0.8 threshold. This phenomenon is rarely observed in such evaluations [49], which suggests that participants assessed their interactions with Harmony positively throughout the study. While the score range remains mostly similar across sessions, the scales receiving the minimum and maximum scores differ: *Visual aesthetics* (M=1.25, SD=1.02) and *Attractiveness* (M=2.18, SD=0.71) define the range in S3 whereas *Adaptability* (M=1.32, SD=1.39), *Attractiveness* (M=2.18, SD=0.80), and *Stimulation*

[3] A Mars-day, or a *Sol*, constitutes a solar day on Mars, *i.e.*, the apparent interval between two successive returns of the Sun to the same meridian as seen by an observer on Mars, which is approximately 24 hours, 39 minutes, and 35 seconds on Earth.

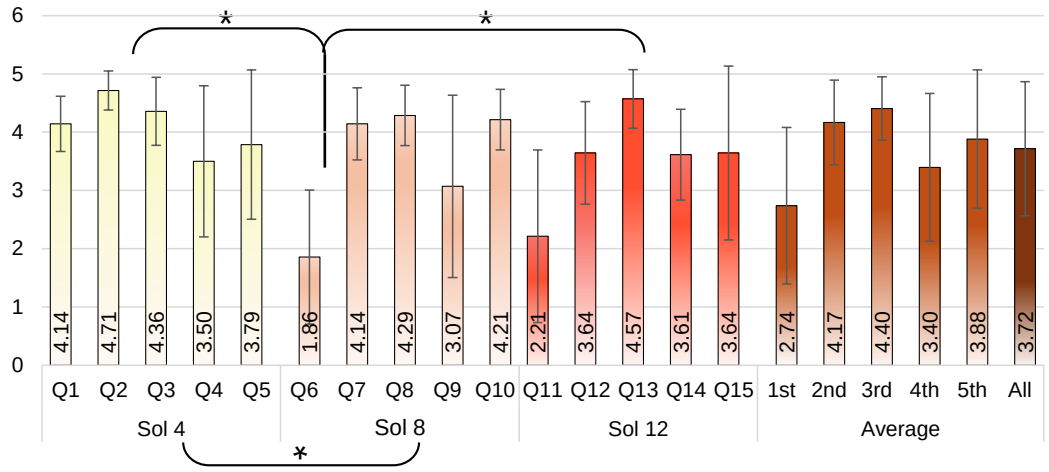

Figure 7 Quality scores of the answers provided by Harmony, from 0 (failure) to 1 (lowest quality) to 5 (highest quality). Error bars show 95% CIs; significance levels are reported at p=.05*.

(M=2.18, SD=0.66) in S4. The following trends can be distinguished for the observed UX:

- *V-shaped curves* undergo a noticeable drop, represented by a local minimum of the SMS starting in the first session on Mars, then gradually rising to S4; see *Perspicuity*, *Trust*, *Value*, and *Intuitive use* ($\frac{4}{13}$=31%). For example, *Value* started from a mean score of M=1.82 (SD=0.76) in S2 and reached M=1.79 (SD=0.86) in S4.
- *Inverted V-shaped curves* represent the opposite trend with a local maximum and ending with a lower score; see *Efficiency* ($\frac{1}{13}$=8%).
- *Overall upward trends* progressively increase from the first to the last evaluation session; see *Attractiveness*, *Stimulation*, and *Visual aesthetics* ($\frac{3}{13}$=23%). For example, *Stimulation* increased from S2 (M=1.64, SD=1.04) to S4 (M=2.18, SD=0.66).
- *Overall downward trends* progressively decrease from the first to the last evaluation session; see *Dependability*, *Novelty*, *Adaptability*, *Usefulness*, and *Trustworthiness* ($\frac{5}{13}$=38%).

These results should be interpreted based on the importance attributed by the participants, measured with the Scale-Mean-Importance (SMI) scores, as detailed below:

- *V-shaped curves* ($\frac{4}{13}$=31%) for *Value*, *Visual Aesthetics*, *Intuitive use*, and *Trustworthiness*.
- *Inverted V-shaped curves* ($\frac{5}{13}$=38%) for *Efficiency*, *Perspicuity*, *Dependability*, *Adaptability*, and *Novelty*.
- *Upward curves* ($\frac{3}{13}$=23%) for *Attractiveness*, *Stimulation*, and *Trust*.
- *Downward curve* ($\frac{1}{13}$=8%) for *Usefulness*.

5.2 Perceptions of Harmony's Answers

Figure 7 shows Quality-Score evaluations of the answers provided by Harmony. A one-way ANOVA revealed a statistically significant difference across the conditions ($F_{14,90}$=2.22, *p=.013) with a large effect size (η^2=0.26). Tukey's HSD Test for multiple comparisons found that the mean value of Quality-Score was significantly different between Q_2 and Q_6 (p=.027, 95% CI=[0.16, 5.55]) and between Q_6 and Q_{13} (p=.046, 95% CI=[0.02, 5.40]). Wilcoxon signed-rank tests among the group of five questions in each session (Sol 4, Sol 8, and Sol 12) revealed Sol 4 and Sol 8 significantly different (z=1.90, p=.028, r=.22).

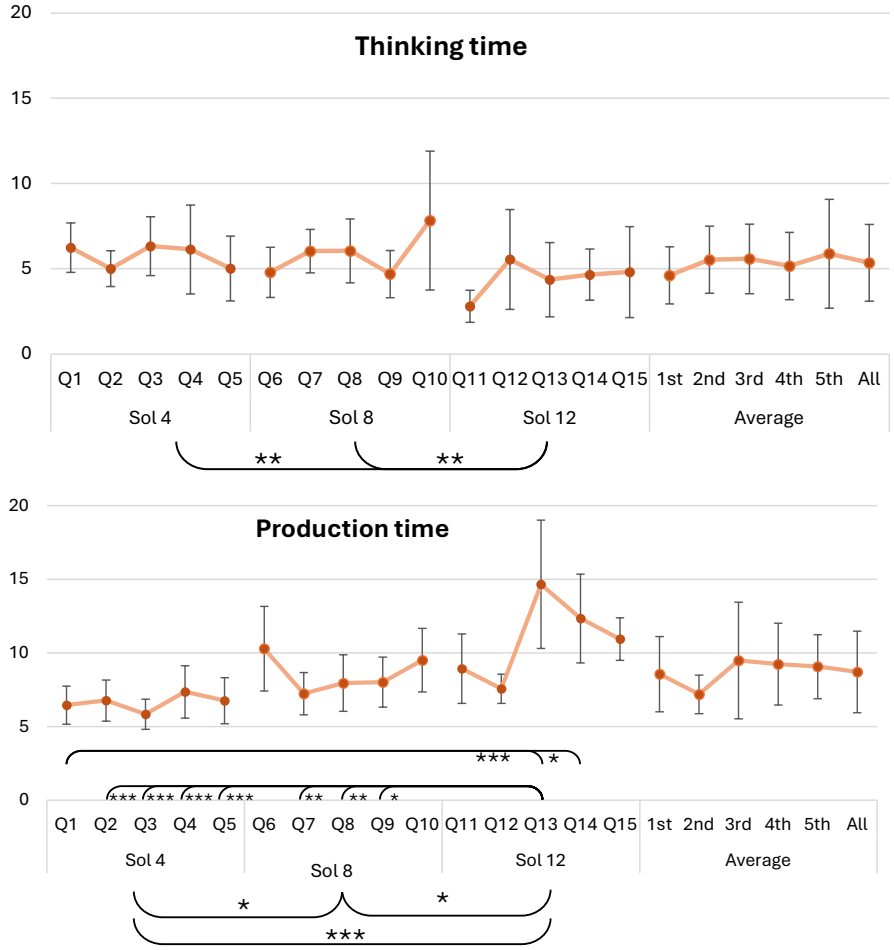

Figure 8 Thinking time (top) and production time (bottom) of the questions asked by the analog astronauts. Error bars show 95% CIs; significance levels are $p \leq .05^{*}$, $p \leq .01^{**}$, and $p \leq .001^{***}$.

5.3 User Performance

Figure 8, top shows the average THINKING-TIME participants needed for the various tasks, with no statistically significant difference ($F_{14,90}$=1.01, p=.056, *n.s.*). However, when we considered the total THINKING-TIME per session, a Wilcoxon signed-rank test revealed that the Sol 4 and Sol 12 conditions were significantly different (z=2.58, p=.0049) with a medium effect size (r=.31), as well as Sol 8 and Sol 12 (z=2.62, p=.0043). This finding suggests that the participants benefited from a learning effect. We did not find any other significant difference between the other sessions and between the five orders of questions, suggesting that the participants needed a similar time to address questions. Figure 8, right shows the average PRODUCTION-TIME participants needed for the various tasks. A one-way ANOVA revealed a statistically significant difference across the conditions ($F_{14,90}$=4.32, $p \leq .001$) with a medium effect size (η^2=.40). The average production time significantly increased from one session to the next ($F_{2,102}$=14.11, $p \leq .001$) with a small effect size (η^2=.22). This result suggests that the participants were progressively more careful in how they entered the vocal command to implement the interaction with Harmony.

Figure 9, left shows how many trials were performed for each question: on average, the questions were assessed as satisfying after the first trial (71%), the second trial (14%), and

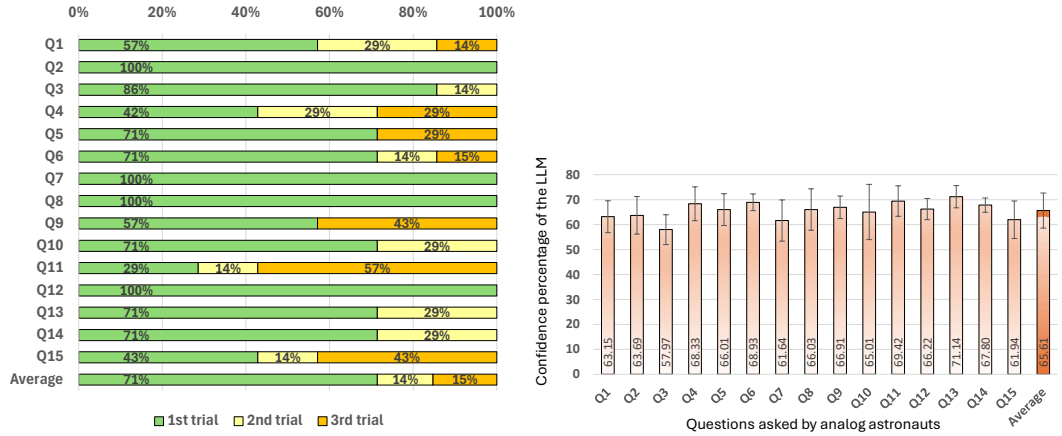

Figure 9 Trials per question (left, see Section 4.3) and CONFIDENCE levels (right).

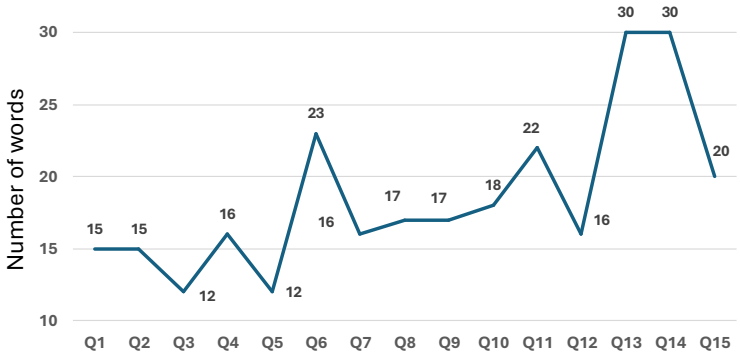

Figure 10 Number of words per question (Q_1 to Q_{15}).

the third trial (15%), respectively. The answers provided by Harmony to Q_2, Q_7, Q_8, and Q_{12} were judged satisfactory during the first trial mainly due to our participants' familiarity with the question domain (*e.g.*, food, management). Figure 9, right shows the CONFIDENCE evaluations of the answers provided by Harmony, ranging from a minimum of 57.97% for Q_3 to a maximum of 71.14% for Q_{13}. The mean of 65.61% suggests that the participants had moderate confidence in the answers provided by Harmony, probably due to a lack of traceability and explanation, in line with the TRUST scores. Participants repeatedly expressed doubts about the system's answers, *e.g.*, Q_{11} about water purification and Q_{15} in medicine, except in the case of participants whose area of expertise was relevant. Figure 10 shows the average number of words per question with Q_{13} and Q_{14}, Q_3 and Q_5 standing out with the highest (30) and the lowest number of words (12), respectively.

5.4 Summary

Table 2 provides a summary of the main UX trends identified in our experiment:

- *Attractiveness*: The value and importance associated with this pragmatic dimension increased over the evaluation sessions, suggesting that participants developed a more favorable overall impression and came to recognize the significance of this dimension.
- *Efficiency*: This pragmatic dimension was the most affected by the ICE conditions, which was an expected result since performance represents the primary criterion in such

Table 2 Evolution of UX scales over sessions: $\simeq$= ratio similar to Earth, $<$ = ratio inferior to Earth, $>$ = ratio superior to Earth, $\ll$ = ratio largely inferior to Earth, and $\gg$ = ratio largely superior to Earth. SMS = scale mean score, SMI = scale mean importance.

Scale	SMS	SMI	Scale	SMS	SMI
Attractiveness	$\gg$	$\gg$	*Trust*	$\gg$	$\gg$
Efficiency	$\ll$	$\simeq$	*Usefulness*	$\ll$	$\ll$
Perspicuity	$<$	$<$	*Value*	$\simeq$	$\gg$
Dependability	$\ll$	$\ll$	*Visual Aesthetics*	$>$	$\simeq$
Stimulation	$\gg$	$\gg$	*Intuitive Use*	$\simeq$	$>$
Novelty	$\ll$	$<$	*Trustworthiness*	$<$	$<$
Adaptability	$\ll$	$\ll$			

contexts [33, 34, 35], while its importance remained consistent across sessions.

- *Perspicuity*: This dimension showed a decline in both value and importance across the sessions, though not significantly. This finding suggests that further attention is needed to ensure transparency of system functionality for increased learnability and ease of use.
- *Dependability*: This pragmatic dimension revealed the greatest decline in value after *Efficiency,* including in terms of participants' perceived importance. Moreover, participants' sense of control over the LLM decreased over time, likely due to the similarity of the responses received and the lack of variation in the level of detail.
- *Stimulation*: This hedonic dimension revealed a growing importance surpassing what is typically observed in a conventional Earth-like setting. Stimulation was strongly affected from the beginning to the end, where participants reported feeling increasingly stimulated when using the LLM, due to the growing importance also felt in terms of *Value*.
- *Novelty*: The scores of this dimension decreased progressively across the evaluation sessions as participants become more accustomed to Harmony, resulting in a reduction in both its perceived value and importance.
- *Adaptability*: While participants appreciated the consistency of the answers, they expressed concerns that Harmony did not provide any means to adapt responses to their individual needs, preferences, or level of expertise. For less experienced users, responses could benefit from progressively increasing levels of detail according to the request, whereas the more experienced participants preferred concise summaries. These findings indicate a clear need for adaptive response mechanisms according to the context of the question, including its urgent or safety-critical nature.
- *Trust*: Participants reported feeling increasingly secure when using the system, mainly due to the quality of the responses. Confidence was also reinforced by the reduced number of trials, an observation consistent with findings reported in previous work [43].
- *Usefulness*: This dimension, initially recognized for its benefits, declines over the evaluation sessions, a finding that highlights the need for strategies to maintain perceived usefulness over time.
- *Visual aesthetics*: This dimension became increasingly valued over the sessions with a stable level of perceived importance. As such, it does not need major changes.
- *Intuitive use*: This dimension remained constant, but warrants further attention due to its growing perceived importance.
- *Trustworthiness*: This dimension requires increased attention as participants perceived it as deteriorating over sessions, casting doubt on Harmony's interaction capabilities.

The dimensions requiring significant improvement are *Efficiency*, *Dependability*, *Adaptability*, *Usefulness*, and *Trustworthiness* (the latter to a lesser extent since it is somewhat compensated by *Trust*). In contrast, *Attractiveness*, *Stimulation*, *Trust*, and *Visual Aesthetics* showed positive development over time, despite the ICE conditions becoming more constraining, and do not warrant immediate action. Lastly, *Perspicuity* and *Trustworthiness* could be slightly improved, whereas *Value* and *Intuitive Use* received consistently positive evaluations.

6 Conclusion and Future Work

We reported results from an experiment involving seven analog astronauts who evaluated their user experience of interacting with Harmony, an LLM designed for real-time technical assistance in ICE environments. We identified the UX dimensions that require improvement, others necessitating optional enhancements, while others were consistently rated positively across multiple evaluations. In this context, LLMs have the potential to revolutionize astronaut assistance by providing intelligent, context-aware support during space missions. While technical challenges remain, continued research and development have the potential to significantly enhance mission efficiency, safety, and success. For example, designing an LLM that supports mental well-being in ICE environments, *e.g.*, through emotional intelligence and personalized interactions, should help enhance *Efficiency* and *Adaptability* through personalized LLM behavior and an adapted tone for long-term engagement. The LLM's responses should be transparent and easy to understand for more trusting and rewarding human-AI collaboration, which requires sensing the context of use and employing techniques to optimize LLM processing. Moreover, collaboration between humans and AI represents a critical step toward sustainable and autonomous space exploration. Future work can compare various adaptation techniques to address the UX dimensions necessitating improvement, particularly by considering alternative interaction modalities [58]. In our experiment, we tested only a graphical user interface with voice-based input, excluding other interaction modalities, such as gesture commands for hands-free operation and haptic feedback when astronauts wear gloves [32], which can be explored in future work.

Acknowledgments

Jean Vanderdonckt and Radu-Daniel Vatavu acknowledge support from a grant of the Romanian Ministry of Research, Innovation and Digitization, CCCDI-UEFISCDI, PN-IV-P8-8.3-PM-RO-BE-2024-0003, within PNCDI IV and Wallonnie-Bruxelles International (WBI) Grant no. 650763 (WBI ref. PAD/CC/VP/TM Roumanie - SUB/2024/650763)—XXR (Novel Extended Reality Models and Software Framework for Interactive Environments with Extreme Conditions). Jean Vanderdonckt is also supported by the EU EIC Pathfinder-Awareness Inside challenge Symbiotik project (1 Oct. 2022-30 Sept. 2026) under Grant no. 101071147. We thank all the analog astronauts of the MARS-UCLouvain Crew 296 "Atlas". Also see a video at `https://www.instagram.com/share/BBJLhHqulE`.

References

1 Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. Guidelines for Human-AI Interaction. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, CHI '19, page 1–13, New York, NY, USA, 2019. ACM. `doi:10.1145/3290605.3300233`.

2 Paul T. Bartone, Gerald P. Krueger, and Jocelyn V. Bartone. Individual differences in adaptability to isolated, confined, and extreme environments. *Aerospace Medicine and Human Performance*, 89(6):536–546, June 2018. URL: https://www.doi.org/10.3357/AMHP.4951.2018, doi:10.3357/AMHP.4951.2018.

3 Leonie Bensch, Tommy Nilsson, Paul de Medeiros, Florian Dufresne, Andreas Gerndt, Flavie Rometsch, Georgia Albuquerque, Frank Flemisch, Oliver Bensch, Michael Preutenborbeck, and Aidan Cowley. Towards balanced astronaut-oriented design for future eva space technologies. In *Proceedings of the third International Conference on Human-Computer Interaction for Space Exploration*, SpaceCHI 3.0, Boston, MA, USA, 2023. MIT Media Lab. URL: https://elib.dlr.de/201717.

4 Leonie Bensch, Tommy Nilsson, Jan Wulkop, Paul de Medeiros, Nicolas Daniel Herzberger, Michael Preutenborbeck, Andreas Gerndt, Frank Flemisch, Florian Dufresne, Georgia Albuquerque, and Aidan Cowley. Designing for Human Operations on the Moon: Challenges and Opportunities of Navigational HUD Interfaces. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, pages 718:1–718:21. ACM, 2024. doi:10.1145/3613904.3642859.

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. *CoRR*, abs/2409.14206, 2024. URL: https://doi.org/10.48550/arXiv.2409.14206.

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

7 Angelica M. Bonilla Fominaya, Rong Kang Chew, Matthew L. Komar, Jeremia Lo, Alexandra Slabakis, Ningjing Sun, Yunyi Zhang, and David Lindlbauer. MoonBuddy: A Voice-Based Augmented Reality User Interface That Supports Astronauts During Extravehicular Activities. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22 Adjunct, New York, NY, USA, 2022. ACM. doi:10.1145/3526114.3558690.

8 Ruijia Cheng, Alison Smith-Renner, Ke Zhang, Joel R. Tetreault, and Alejandro Jaimes. Mapping the design space of Human-AI interaction in text summarization. *CoRR*, abs/2206.14863, 2022. URL: https://doi.org/10.48550/arXiv.2206.14863.

9 William J Clancey. Participant Observation of a Mars Surface Habitat Mission Simulation. *Habitation*, 11(1):27–47, 2006. doi:10.3727/154296606779507132.

10 Joëlle Coutaz, James L. Crowley, Simon Dobson, and David Garlan. Context is key. *Communications of the ACM*, 48(3):49–53, mar 2005. doi:10.1145/1047671.1047703.

11 Matthew Deans, Jessica J. Márquez, Tamar Cohen, Matthew J. Miller, Ivonne Deliz, Steven Hillenius, Jeffrey Hoffman, Yeon Jin Lee, David Lees, Johannes Norheim, and Darlene S. S. Lim. Minerva: User-centered science operations software capability for future human exploration. In *Proceedings of the IEEE Aerospace Conference*, AERO '17, pages 1–13, 2017. doi:10.1109/AERO.2017.7943609.

12 Florian Dufresne, Tommy Nilsson, Geoffrey Gorisse, Enrico Guerra, André Zenner, Olivier Christmann, Leonie Bensch, Nikolai Anton Callus, and Aidan Cowley. Touching the Moon: Leveraging Passive Haptics, Embodiment and Presence for Operational Assessments in Virtual Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24. ACM, 2024. URL: https://doi.org/10.1145/3613904.3642292.

13 Ariel Ekblaw, Juliana Cherston, Fangzheng Liu, Irmandy Wicaksono, Don Derek Haddad, Valentina Sumini, and Joseph A. Paradiso. From UbiComp to Universe-Moving Pervasive Computing Research Into Space Applications. *IEEE Pervasive Computing*, 22(2):27–42, 2023. doi:10.1109/MPRV.2023.3242667.

14 Julian Barrera Esquinas, Airbus Sas, and Rond-Point Maurice Bellonte. Evolution of Human-Device Interface in the Field of Technical Documentation. In R. John Hansman and Stéphane Chatty, editors, *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics*, HCI-Aero '02, Washington, DC, USA, 2002. American Association for Artificial Intelligence. URL: https://aaai.org/papers/hci02-013-evolution-of-human-device-interface-in-the-field-of-technical-documentation.

15 Oscar Firschein. Artificial Intelligence for Space Station Automation: Crew Safety, Productivity, Autonomy, Augumented Capability, 1986. URL: https://api.semanticscholar.org/CorpusID:58730479.

16 Sands Fish. How to Design Interplanetary Apps. Medium, 2018. URL: https://sandsfish.medium.com/how-to-design-interplanetary-apps-22ebefec097d.

17 Sands Fish. Orientation-Responsive Displays for Microgravity. In *Proceedings of the SpaceCHI 2.0 Workshop, Advancing Human-Computer Interaction for Space Exploration at CHI 2022*, SpaceCHI 2.0, MA, USA, 2022. MIT Media Lab. URL: https://drive.google.com/open?id=1BUQDkbt6tSCJ759Z00FuaEmm6Af5h2z9.

18 B.H. Foing, C. Stoker, J. Zavaleta, P. Ehrenfreund, C. Thiel, P. Sarrazin, D. Blake, J. Page, V. Pletser, J. Hendrikse, and et al. Field astrobiology research in Moon–Mars analogue environments: instruments and methods. *International Journal of Astrobiology*, 10(3):141–160, 2011. doi:10.1017/S1473550411000036.

19 Ana Rita Goncalves Freitas, Alexander Schülke, Simon Glaser, Pitt Michelmann, Thanh Nguyen Chi, Lisa Marie Schröder, Zahra Fadavi, Gaurav Talekar, Jette Ternieten, Akash Trivedi, Jana Wahls, Warda Masood, Christiane Heinicke, and Johannes Schöning. Conversational User Interfaces to support Astronauts in Extraterrestrial Habitats. In *Proceedings of the 20th International Conference on Mobile and Ubiquitous Multimedia*, MUM '21, page 169–178, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3490632.3490673.

20 Gregory Goth. Software on Mars. *Commun. ACM*, 55(11):13–15, nov 2012. doi:10.1145/2366316.2366321.

21 Carsten Hartmann, Franca Speth, Dieter Sabath, and Florian Sellmaier. METIS: An AI Assistant Enabling Autonomous Spacecraft Operations for Human Exploration Missions. In *Proceedings of the 2024 IEEE Aerospace Conference*, pages 1–22, 2024. doi:10.1109/AERO58975.2024.10521154.

22 Marc Hassenzahl, Sarah Diefenbach, and Anja Göritz. Needs, affect, and interactive products – facets of user experience. *Interacting with Computers*, 22(5):353–362, 2010. Modelling user experience - An agenda for research and practice. doi:10.1016/j.intcom.2010.04.002.

23 Sandra Häuplik-Meusburger and Sheryl Bishop. *Introduction to ICE*, pages 1–8. Space and Society. Springer International Publishing, Cham, 2021. doi:10.1007/978-3-030-69740-2_1.

24 Andreas Hinderks, Dominique Winter, Martin Schrepp, and Jörg Thomaschewski. Applicability of user experience and usability questionnaires. *J. Univers. Comput. Sci.*, 25(13):1717–1735, 2019. URL: http://www.jucs.org/jucs_25_13/applicability_of_user_experience.

25 Lauren Blackwell Landon, Grace L. Douglas, Meghan E. Downs, Maya R. Greene, Alexandra M. Whitmire, Sara R. Zwart, and Peter G. Roma. The Behavioral Biology of Teams: Multidisciplinary Contributions to Social Dynamics in Isolated, Confined, and Extreme Environments. *Frontiers in Psychology*, 10, 2019. doi:10.3389/fpsyg.2019.02571.

26 Lauren Blackwell Landon, Jessica J. Márquez, and Eduardo Salas. uman Factors in Spaceflight: New Progress on a Long Journey. *Human Factors*, 65(6):973–976, 2023. doi:10.1177/00187208231170276.

27 John Leach. Psychological factors in exceptional, extreme and torturous environments. *Extreme Physiology & Medicine*, 5(7), 2016. doi:10.1186/s13728-016-0048-y.

28 Lauren B. Leveton, Camille Shea, Kelley J. Slack, Kathryn E. Keeton, and Lawrence A. Palinkas. Antarctica Meta-analysis: Psychosocial Factors Related to Long-duration Isolation and

Confinement. In *Proceedings of Human Research Program Investigators Workshop*. Universities Space Research Association, 2009. doi:https://ntrs.nasa.gov/citations/20090007551.

29 Shu-Yu Lin and Katya Arquilla. Quantifying proprioceptive experience in microgravity. In *Proceedings of the SpaceCHI 2.0 Workshop, Advancing Human-Computer Interaction for Space Exploration at CHI 2022*, SpaceCHI 2.0, MA, USA, 2022. MIT Media Lab. URL: https://drive.google.com/open?id=14adCKB1U5m2-0kiL9BL02rD6-uUkoavd.

30 Rhema Linder, Chase Hunter, Jacob McLemore, Senjuti Dutta, Fatema Akbar, Ted Grover, Thomas Breideband, Judith W. Borghouts, Yuwen Lu, Gloria Mark, Austin Z. Henley, and Alex C. Williams. Characterizing Work-Life for Information Work on Mars: A Design Fiction for the New Future of Work on Earth. *Proceedings of ACM Human-Computer Interaction*, 6(GROUP), jan 2022. doi:10.1145/3492859.

31 Julie Manon, Vladimir Pletser, Michael Saint-Guillain, Jean Vanderdonckt, Cyril Wain, Jean Jacobs, Audrey Comein, Sirga Drouet, Julien Meert, Ignacio Jose Sanchez Casla, Olivier Cartiaux, and Olivier Cornu. An Easy-To-Use External Fixator for All Hostile Environments, from Space to War Medicine: Is It Meant for Everyone's Hands? *Journal of Clinical Medicine*, 12(14), 2023. doi:10.3390/jcm12144764.

32 Julie Manon, Jean Vanderdonckt, Michael Saint-Guillain, Vladimir Pletser, Cyril Wain, Jean Jacobs, Audrey Comein, Sirga Drouet, Julien Meert, Ignacio Sanchez Casla, Olivier Cartiaux, and Olivier Cornu. A Multi-Session Evaluation of a Haptic Device in Normal and Critical Conditions: a Mars Analog Mission. *International Journal of Interactive Multimedia and Artificial Intelligence*, 9(3):164–174, 06 2025. URL: https://www.ijimai.org/journal/bibcite/reference/3579, doi:10.9781/ijimai.2025.04.001.

33 Jessica J. Márquez and Mary L. Cummings. Design and evaluation of path planning decision support for planetary surface exploration. *Journal of Aerospace Computing, Information, and Communication*, 5(3):57–71, 2008. doi:10.2514/1.26248.

34 Jessica J. Márquez, Tamsyn Edwards, John A. Karasinski, Candice N. Lee, Megan C. Shyr, Casey L. Miller, and Summer L. Brandt. Human performance of novice schedulers for complex spaceflight operations timelines. *Human Factors*, 65(6):1183–1198, 2023. doi:10.1177/00187208211058913.

35 Jessica J. Márquez, Lauren Blackwell Landon, and Eduardo Salas. The next giant leap for space human factors: The opportunities. *Human Factors*, 65(6):1279–1288, 2023. doi:10.1177/00187208231174955.

36 Kaitlin R. McTigue, Megan E. Parisi, Tina L. Panontin, Shu-Chieh Wu, and Alonso H. Vera. How to keep your space vehicle alive: Maintainability design principles for deep-space missions. In *Proceedings of SpaceCHI 3.0, A Conference on Human-Computer Interaction for Space Exploration*, SpaceCHI 3.0, MA, USA, 2023. MIT Media Lab. URL: https://human-factors.arc.nasa.gov/publications/SpaceCHI2023_Maintainability.pdf.

37 Anna-Lena Meiners, Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. A Benchmark for the UEQ+ Framework: Construction of a Simple Tool to Quickly Interpret UEQ+ KPIs. *Int. J. Interact. Multim. Artif. Intell.*, 9(1):104, 2024. doi:10.9781/IJIMAI.2023.05.003.

38 Tommy Nilsson, Leonie Bensch, Florian Dufresne, Flavie Rometsch, Paul de Medeiros, Enrico Guerra, Florian Saling, Andrea Casini, and Aidan Cowley. Out of this world design: Bridging the gap between space systems engineering and participatory design practices. In *Proceedings of SpaceCHI 3.0, A Conference on Human-Computer Interaction for Space Exploration*, SpaceCHI 3.0, MA, USA, 2023. MIT Media Lab. URL: https://spacechi.media.mit.edu/spacechi-2023-program.

39 Tommy Nilsson, Flavie Rometsch, Leonie Becker, Florian Dufresne, Paul Demedeiros, Enrico Guerra, Andrea Emanuele Maria Casini, Anna Vock, Florian Gaeremynck, and Aidan Cowley. Using Virtual Reality to Shape Humanity's Return to the Moon: Key Takeaways from a Design Study. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. ACM. doi:10.1145/3544548.3580718.

40 Marianna Obrist, Yunwen Tu, Lining Yao, and Carlos Velasco. Space food experiences: Designing passenger's eating experiences for future space travel scenarios. *Frontiers in Computer Science*, 1, 2019. doi:10.3389/fcomp.2019.00003.

41 Lawrence A. Palinkas and Peter Suedfeld. Psychosocial issues in isolated and confined extreme environments. *Neuroscience & Biobehavioral Reviews*, 126:413–429, 2021. doi:10.1016/j.neubiorev.2021.03.032.

42 Piyush Pant, Anand Singh Rajawat, S.B. Goyal, Amol Potgantwar, Pradeep Bedi, Maria Simona Raboaca, Neagu Bogdan Constantin, and Chaman Verma. AI based Technologies for International Space Station and Space Data. In *Proceedings of 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*, pages 19–25, 2022. doi:10.1109/SMART55829.2022.10046956.

43 Kanak Parmar and Nathan L. Parrish. Assurance of human-ai interaction based systems for spaceflight: A discussion of critical aspects to increase mutual trust and reliability. In *AIAA SCITECH 2024 Forum*. American Institute of Aeronautics and Astronautics, 2024. doi:10.2514/6.2024-2063.

44 Pat Pataranutaporn, Valentina Sumini, Ariel Ekblaw, Melodie Yashar, Sandra Häuplik-Meusburger, Susanna Testa, Marianna Obrist, Dorit Donoviel, Joseph Paradiso, and Pattie Maes. Spacechi: Designing human-computer interaction systems for space exploration. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, New York, NY, USA, 2021. ACM. doi:10.1145/3411763.3441358.

45 Pat Pataranutaporn, Valentina Sumini, Melodie Yashar, Susanna Testa, Marianna Obrist, Scott Davidoff, Amber M. Paul, Dorit Donoviel, Jimmy Wu, Sands A Fish, Ariel Ekblaw, Albrecht Schmidt, Joe Paradiso, and Pattie Maes. SpaceCHI 2.0: Advancing Human-Computer Interaction Systems for Space Exploration. In *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems*, CHI EA '22, New York, NY, USA, 2022. ACM. doi:10.1145/3491101.3503708.

46 Vladimir Pletser and Bernard Foing. European Contribution to Human Aspect Investigations for Future Planetary Habitat Definition Studies: Field Tests at MDRS on Crew Time Utilisation and Habitat Interfaces. *Microgravity Science and Technology*, 23:199–214, 2011. doi:10.1007/s12217-010-9251-4.

47 Antônio Victor Figueiredo Porto and Maria Elizabeth Sucupira Furtado. Framework to specify dialogues for natural interaction with conversational assistants applied in prompt engineering. In Kohei Arai, editor, *Intelligent Systems and Applications*, pages 231–253, Cham, 2024. Springer Nature Switzerland.

48 Michael Saint-Guillain, Jean Vanderdonckt, Nicolas Burny, Vladimir Pletser, Tiago Vaquero, Steve Chien, Alexander Karl, Jessica Márquez, Cyril Wain, Audrey Comein, Ignacio S. Casla, Jean Jacobs, Julien Meert, Cheyenne Chamart, Sirga Drouet, and Julie Manon. Enabling astronaut self-scheduling using a robust advanced modelling and scheduling system: An assessment during a Mars analogue mission. *Advances in Space Research*, 72(4):1378–1398, 2023. doi:10.1016/j.asr.2023.03.045.

49 Martin Schrepp. Measuring user experience with modular questionnaires. In *Proceedings of International Conference on Advanced Computer Science and Information Systems*, ICACSIS '21, pages 1–6, Piscataway, NJ, USA, 2021. IEEE Press. doi:10.1109/ICACSIS53237.2021.9631321.

50 Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *Int. J. of Interactive Multimedia and Artificial Intelligence.*, 4(4):40–44, 2017. doi:10.9781/ijimai.2017.445.

51 Martin Schrepp and Jörg Thomaschewski. Design and Validation of a Framework for the Creation of User Experience Questionnaires. *Int. J. of Interactive Multimedia and Artificial Intelligence*, 5(7):88–95, 2019. doi:10.9781/IJIMAI.2019.06.006.

52 Shivang Shelat, John A. Karasinski, Erin E. Flynn-Evans, and Jessica J. Márquez. Evaluation of User Experience of Self-scheduling Software for Astronauts: Defining a Satisfaction Baseline. In

Don Harris and Wen-Chin Li, editors, *Engineering Psychology and Cognitive Ergonomics*, pages 433–445, Cham, 2022. Springer International Publishing. doi:10.1007/978-3-031-06086-1_34.

53 Shivang Shelat, Jessica J. Márquez, Jimin Zheng, and John A. Karasinski. Collaborative System Usability in Spaceflight Analog Environments through Remote Observations. *Applied Sciences*, 14(5), 2024. doi:10.3390/app14052005.

54 Peter Suedfeld. Applying Positive Psychology in the Study of Extreme Environments. *Journal of Human Performance in Extreme Environments*, 6(1), 2021. doi:10.7771/2327-2937.1020.

55 Konstantinos Tsiakas and Dave Murray-Rust. Unpacking Human-AI interactions: From Interaction Primitives to a Design Space. *ACM Trans. Interact. Intell. Syst.*, 14(3), August 2024. doi:10.1145/3664522.

56 Martine Van Puyvelde, Daisy Gijbels, Thomas Van Caelenberg, Nathan Smith, Loredana Bessone, Susan Buckle-Charlesworth, and Nathalie Pattyn. Living on the edge: How to prepare for it? *Frontiers in Neuroergonomics*, 3, 2022. doi:10.3389/fnrgo.2022.1007774.

57 Jean Vanderdonckt, Gaëlle Calvary, Joëlle Coutaz, and Adrian Stanciulescu. *Multimodality for Plastic User Interfaces: Models, Methods, and Principles*, pages 61–84. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. doi:10.1007/978-3-540-78345-9_4.

58 Jean Vanderdonckt, Radu-Daniel Vatavu, Julie Manon, Romain Maddox, Michael Saint-Guillain, Philippe Lefevre, and Jessica J. Márquez. UX, but on Mars: Exploring User Experience in Extreme Environments with Insights from a Mars Analog Mission. In *Proceedings of the ACM International Conference on Designing Interactive Systems Conference*, DIS '25, New York, NY, USA, 2025. Association for Computing Machinery. doi:10.1145/3715336.3735706.

59 Jean Vanderdonckt, Radu-Daniel Vatavu, Julie Manon, Michael Saint-Guillain, Philippe Lefevre, and Jessica J. Márquez. Might as Well Be on Mars: Insights on the Extraterrestrial Applicability of Interaction Design Frameworks from Earth. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '24, New York, NY, USA, 2024. Association for Computing Machinery. doi:10.1145/3613905.3650807.

60 Radu-Daniel Vatavu, Jean Vanderdonckt, Julie Manon, Michael Saint-Guillain, Philippe Lefevre, Romain Maddox, and Jessica J. Márquez. Conducting human-computer interaction scientific experiments in extreme environments: Insights from analog mars missions. *Romanian Journal of Information Science and Technology*, 2025.

61 Bernhard Weber and Martin Stelzer. Sensorimotor impairments during spaceflight: Trigger mechanisms and haptic assistance. *Frontiers in Neuroergonomics*, 3, 2022. doi:10.3389/fnrgo.2022.959894.

62 Jimin Zheng, Shivang M. Shelat, and Jessica J. Márquez. Facilitating Crew-Computer Collaboration During Mixed-Initiative Space Mission Planning. In *Proceedings of SpaceCHI 3.0, A Conference on Human-Computer Interaction for Space Exploration*, SpaceCHI 3.0, MA, USA, 2023. MIT Media Lab. URL: https://ntrs.nasa.gov/citations/20230008619.

63 Pierpaolo Zivi, Luigi De Gennaro, and Fabio Ferlazzo. Sleep in Isolated, Confined, and Extreme (ICE): A Review on the Different Factors Affecting Human Sleep in ICE. *Frontiers in Neuroscience*, 14, 2020. doi:10.3389/fnins.2020.00851.