IAC-25-D1.3.4x98736

Future Astronaut-Agent Medical Collaboration Opportunities for Long-Duration Human Spaceflight

Anna B. Wojdecka[a][b]*, Tibor Balint[a], Don Platt[b]

[a] *Royal College of Art, School of Design, Kensington Gore, London SW7 2EU, UK*
[b] *Florida Institute of Technology, 150 W University Blvd, Melbourne, FL 32901, US*
* Corresponding Author anna.wojdecka@network.rca.ac.uk

Abstract

Current medical care of astronauts in Low Earth Orbit (LEO) heavily relies on ground support. However, future long-duration human spaceflight (LDHSF) and exploration-class missions will require novel approaches to design of systems sustaining the health and safety of astronaut crews. Prolonged exposure to the harsh environment will impact every aspect of crews' health, from microgravity-induced changes to human physiology and psychological challenges of living in remote confinement, to cognitive decline due to radiation exposure. Medical decision-making will require optimizing scarce resources alongside prioritizing prevention and early detection.

Leveraging emerging technologies, such as artificial intelligence and biofeedback, along with robotics, presents opportunities for designing novel medical systems that enable astronauts to collaborate with agents as Cyber-Physical-Human (CPH) teams. Although trust and role assignment are well-understood foundations of effective collaboration in healthcare teams and physician-patient relationships on Earth, the astronaut-agent trust-driven synergistic medical collaboration during LDHSF remains largely unexplored.

To address this gap, we adapted a human-centred design approach, developing new research tools for transdisciplinary collaboration involving diverse stakeholders, including space medicine, astronaut and training, human factors, human-centered design, engineering, and human-computer interaction. We conducted qualitative interviews and ran a series of Subject Matter Expert (SME) workshops titled 'Future Health(care) Space Systems: Designing-In Trust into Cyber-Physical Teams and Trustworthy Human-Agent Interactions'.

This paper presents key insights into the challenges and opportunities of CPH trust-driven medical collaboration. We discuss the human-agent shared decision-making and role allocations, encompassing training, prevention, early detection, diagnosis, treatment selection, and care delivery. Lastly, we present transdisciplinary recommendations for designing new CPH-oriented interfaces that capitalize on the strengths of both humans and agents and foster designing-in justifiable trust in the design of future CPH-team-oriented medical systems.

Keywords: Human-Agent Trust, Astronaut-Agent Collaboration, Human-AI Medical Interfaces, Trust-Centered Design

Nomenclature

$T_{H \to A}$	human trust in agent
$T_{A \to H}$	agent trust in human
$T_{A \to A}$	agent self-trust

Acronyms/Abbreviations

CPH	Cyber-Physical-Human
EVA	Extravehicular activity
HCD	human-centered design
HUD	Heads-Up Display
ISS	International Space Station
LEO	Low Earth Orbit
LDHSF	Long-Duration Human Spaceflight
MBSE	Model-Based Systems Engineering
MO	Medical Officer
PoC	Point of Care
SMEs	Subject Matter Experts
SysML	Systems Modeling Language

1. Introduction

Current space medical systems for astronauts at the International Space Station (ISS) are designed around ground-based medical oversight, with protocols for stabilization and contingency evacuation to terrestrial care in off-nominal scenarios [1]. The design of future space medical systems for exploration missions requires a paradigm shift toward comprehensive onboard medical care, maximizing crew autonomy. To minimize risks, the exploration medical systems will need to provide the highest level of care to mitigate the negative impact of the space environment on the human body [2], from prevention and early detection to treatment delivery, including supporting complex clinical decision-making, optimizing resource allocation, guiding on just-in-time (JIT) training and treatment delivery [3,4].

Although leveraging new technologies, including biomonitoring and artificial agents, presents new possibilities for designing systems capable of monitoring, predicting, and supporting both crew health and independent decision-making, it also raises new challenges for system design. Astronauts working with agents as Cyber-Physical-Human (CPH) teams will rely on an interface as a primary channel of communication. Designing interfaces that facilitate effective astronaut–agent interactions thus becomes a central component of CPH-oriented system design.

Effective medical collaboration relies on mutual trust. Just like in the healthcare context, where a trusting doctor–patient relationship supports effective problem-solving and shared decision-making [5], the astronauts-agents interactions need to be designed with mutual trust consideration in mind. Importantly, the systems should only be trusted if they are trustworthy, and the ability to perform the assigned task is a fundamental aspect of trustworthiness [6]. The roles and tasks need to be allocated according to strengths to support trust in astronaut-agent teaming. An agent can offer 24-hour alertness and computational power, and its capabilities are preferred for tasks requiring precision, accuracy, tracking and comparing data, pattern recognition, conducting complex calculations, probabilistic analysis, or prognostic simulations. Humans are unmatched in their creativity in off-nominal situations, adaptability to the unexpected, ethical judgment, and social interaction, while an agent can offer a judgment that is free of emotional bias. Interaction design must leverage these complementary strengths, while also accounting for dynamic conditions—for example, when the system is prone to error or faced with problems outside its scope, or when the astronaut is incapacitated, hypoxic, in pain, or sleep-deprived such that their performance is severely impaired. The interface and interaction design needs to account not only for the human operator's trust in the agent, but also for the agent's ability to recognize when the human is unable to perform the assigned task and requires greater assistance [7,8]. The mutual trust needs to be sustained throughout the interaction, as the demands and tasks shift throughout the interaction.

While scenario-based methods are established in medicine [9], service design and human-centered design [10,11], as well as in Model-Based Systems Engineering (MBSE) [12,13], none explicitly integrate trust dynamics into the interaction. This gap is critical, since mutual, justifiable trust underpins effective human–agent teaming in high-stakes environments. Our research seeks to develop new methods and tools that enable transdisciplinary exploration of trust and its contributing factors in human–agent interaction. By eliciting qualitative insights into trust dynamics and their influence on interaction, we aim to provide a workflow to uncover needs, challenges, and opportunities for interface and system design in exploration medical contexts, supporting the development of CPH-oriented medical system architectures, for synergistic human–agent collaboration.

In this paper, we introduce the trust-aware scenario blueprinting workflow and we illustrate the application of the approach to investigate trust dynamics in astronaut–agent medical collaboration to identify key challenges for effective interaction, and translate trust factor considerations into interface, interaction, and system design needs and opportunities. We further synthesize cross-scenario insights into transdisciplinary design recommendations for fostering effective and trustworthy astronaut–agent collaboration.

2. Methods

This study adopts a human-centered design (HCD) approach with three second-order cybernetic feedback loops [14], employing multidisciplinary stakeholder engagement as a central part of the research—from eliciting insights, developing research methods, synthesis and conceptual development, to contextual validation. In the first phase, the researcher iteratively developed a framework for facilitating transdisciplinary stakeholder collaboration for future contexts [7,15]. In the second phase, following Constructivist Grounded Theory procedures, the researcher conducted qualitative interviews with twenty-five Subject Matter Experts (SMEs) focused on trust in human-agent collaboration, co-developing CPH Trust Maps as tools for systematic consideration of trust in the design of systems for astronaut-agent medical collaboration [7,16]. In the third stage, the researcher ran a series of stakeholder engagement activities, including SME design meetings, collaborative design workshops, and hands-on iterative feedback sessions titled 'Future Health(care) Space Systems: Designing-In Trust into Cyber-Physical Teams and Trustworthy Human-Agent Interactions,' as a part of case study development of a trust-driven astronaut-agent medical interface [8]. This paper focuses on this last research feedback loop, presenting the employed trust-aware scenario blueprinting as a way of eliciting trust-driven needs, and offers transdisciplinary design recommendations synthesized from all the study stages.

2.1 Subjects

The SMEs involved in this study represented a wide range of backgrounds and expertise within the following categories: Space Medicine (including flight surgeons and emergency medicine physicians), Astronauts and Training, Space Systems and Engineering, Architecture and Human Factors, Computing and Human-Computer Interaction. The distribution of the SMEs that contributed to this study included the USA, Canada, the UK, the EU, Australia, New Zealand, India, Mexico, and Brazil.

In-depth qualitative interviews were conducted with twenty-five SMEs. The co-design workshops involved between three and fifteen participants, while the open seminar sessions attracted over fifty attendees.

2.2 Ethics

The study was conducted with ethical clearances: IRB Exemption No. 22-064 and No. 23-005. The first author led the design research study, including the SME interviews, workshops, and SME design meetings.

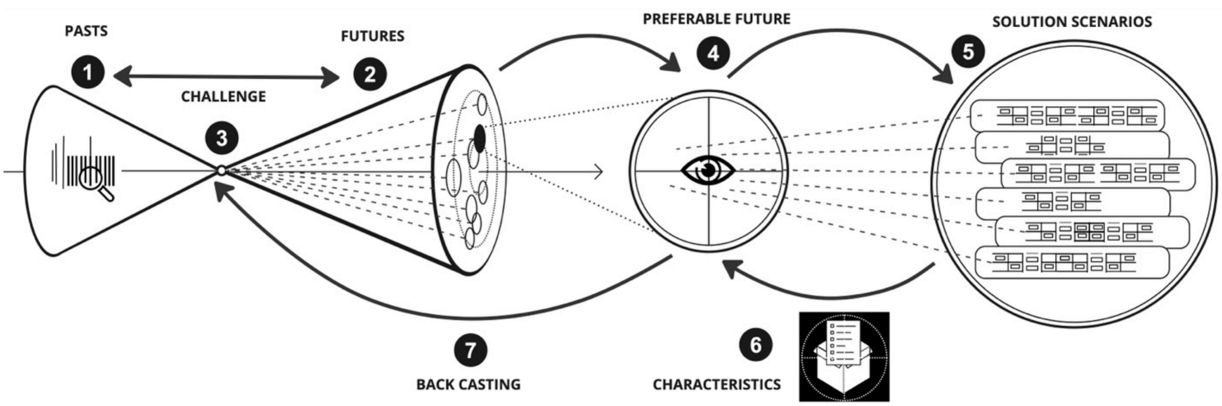

Fig. 1: Framework to Design Out Unwanted Futures [7,15]. Steps involve: (1) Examining the Challenge from Multiple Past Perspectives, (2) Projecting Multiple Futures, (3) Redefining the Challenge, (4) Detailing the Target Future, (5) Generating and Probing Solution Scenarios, (6) Defining the Design Needs/Characteristics, and (7) Hands-On Concept Detailing.

2.2.1 Framework

The study structure followed the Framework for Designing Out Unwanted Futures [7,15], developed to facilitate diverse stakeholder engagement in transdisciplinary projects with long-term time horizons, from co-exploration of pasts, through consideration of preferable future context, envisioning solutions within scenarios, to early concept testing (see Fig. 1). This paper focuses specifically on steps 4–7 of the framework.

2.3 Tools: Trust Maps

As a part of this research, we applied the CPH Trust Maps [7,16] as tools for systematic astronaut-agent trust consideration in designing trust-driven medical systems (see Fig. 2). The maps, derived from qualitative SME interviews, were co-constructed with stakeholders during the earlier stage of the study, corresponding to steps 1-3 of the Future Framework. The CPH Trust Maps codify the derived factors impacting the astronaut↔agent trust relationship: human trust in agent ($T_{H \rightarrow A}$), agent trust in human ($T_{A \rightarrow H}$), and agent self-trust ($T_{A \rightarrow A}$). The factors originating from diverse disciplinary insights are presented together, in relation to the cognitive-affective and personal-mission trust dimensions.

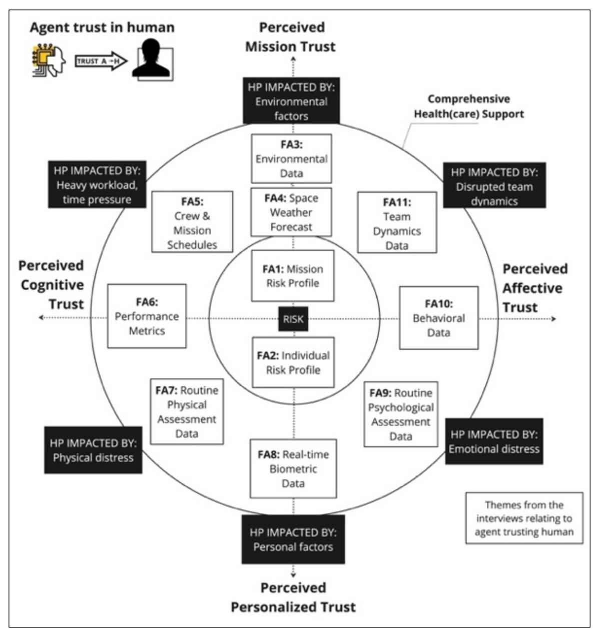

Fig. 2: One of the CPH Trust Maps with trust factors impacting agent trust in human ($T_{A \rightarrow H}$) [7,16].

2.4 Limitations

The initial stages of the study took place during the COVID-19 pandemic, which posed limitations of access to in-person collaborative settings, prompting the adoption of online tools for stakeholder engagement. The interviews and design workshops were conducted both in-person and virtually, utilizing an online whiteboard for real-time stakeholder collaboration.

The highly specialized field of human spaceflight presents challenges of resource availability limitations of globally scarce SMEs, which also implies a challenge in viewpoint diversification within this narrow group of specialists. Our mitigation strategies involved approaching experts during conferences and through the researchers' and stakeholders' space industry networks.

To mitigate potential researcher and disciplinary biases, we employed research triangulation in our approach. We iteratively refined our outputs that arose from the interactions, incorporating reflexivity into our research process. The visual synthesis developed through the iterative process—including conceptual diagrams created by our facilitating researcher—acted as boundary objects [17] in discussions with stakeholders and played a key role in articulating and validating the co-constructed knowledge and insights.

3. Results

This section summarizes the proposed trust-centered scenario-based approach for considering mutual trust in astronaut–agent interaction for medical interface and system design, and is organized into three parts. First, we introduce trust-aware scenario blueprinting, illustrating the workflow and its application in the SME workshops. Next, we discuss the example scenarios generated during stakeholder workshops, highlighting selected challenges and opportunities for trust-driven interface and interaction design. Finally, we synthesize the cross-scenario insights into transdisciplinary recommendations for designing interfaces that capitalize on the strengths of both humans and agents, fostering justifiable trust.

3.1 Trust-Aware Human-Agent Interaction

Fig. 3 illustrates the astronaut-agent interaction facilitated through the interface. From the perspective of medical teaming, the interface is the central part of the system that enables astronaut-agent medical collaboration. The interface design has mutual human↔agent trust implications concerning: human trust in agent ($T_{H\rightarrow A}$), agent trust in human ($T_{A\rightarrow H}$), human self-trust ($T_{H\rightarrow H}$), and agent self-trust ($T_{A\rightarrow A}$) [8]. To collaborate effectively, both the human operator and the agent need to know when to trust each other, and—depending on each party's expertise and ability to perform tasks—how to optimally allocate roles.

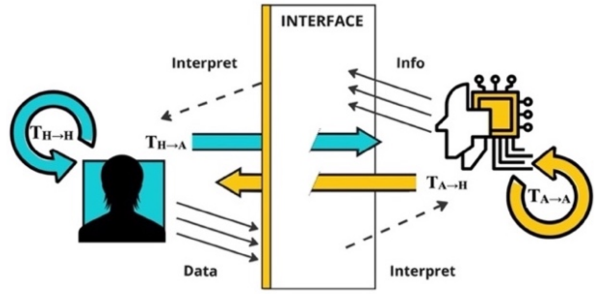

Fig. 3. Human-agent interaction through the interface and trust directions [8].

The diversity of tasks involved in a comprehensive medical care of astronauts on LDHS missions means that the roles might need to be switched as the interaction unfolds, which requires careful trust considerations throughout the interaction. To minimize risks, mutual human↔agent must already be established during pre-mission training. During the mission, trust must be maintained throughout all interactions related to prevention, diagnosis, and emergency medical care.

To consider the mutual human-agent trust and dynamic role allocation throughout health(care) interactions, we proposed a trust-aware interaction blueprinting, which builds on traditional HCD methods involving role-play [18] and service design blueprinting [10,11], incorporating trust as a core interaction component. Fig. 4 shows the developed Trust-Aware Interaction Blueprinting worksheet (populated), which includes a dedicated section for tracking trust dynamics from both a human trust perspective (blue section) and an agent trust perspective (yellow section). Between these sections, at the center of the worksheet, lies the human-agent line of interaction, along which the interaction unfolds. Above the blue astronaut's section, we added the line of crew involvement to specify at which point other actors join the interaction—such as Medical Officer (MO), Point of Care (PoC), or ground medical support (accounting for delay). At the bottom of the agent's section, inspired by the classic blueprint's line of visibility, we added a field related to $T_{A\rightarrow A}$ and the agent's access to resources.

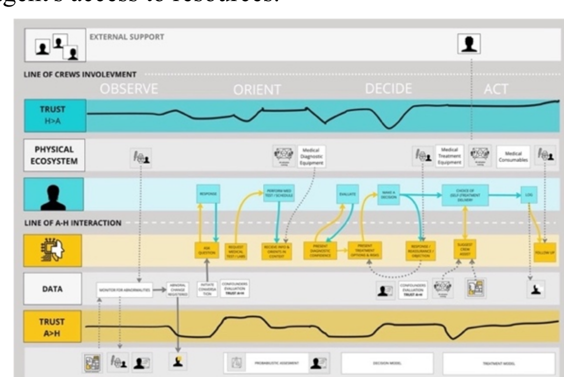

Fig. 4. Trust-aware interaction blueprinting [7,8].

The aim of trust-aware interaction blueprints is to provide an embodied context for envisioning future astronaut–agent collaboration, while uncovering trust-related interaction challenges and opportunities. Analogous to human-centered design user journeys and service design blueprints, the tool functions as a collaborative medium for diverse stakeholders. By enacting roles within envisioned future contexts, HCD role-play and blueprinting support empathy building, the uncovering of needs, and provide a context for ideation. Introducing mutual trust as a central variable emphasizes human–agent collaboration, and brings the trustworthiness of the interaction to the foreground, providing a lens for identifying needs to design for appropriate reliance and effective teamwork.

Our approach explicitly incorporates mutual trust as a core variable. It builds on HCD and service design scenarios, captured through journey maps and blueprints, which provide qualitative focus to exploring experience, human factors, front-stage, and back-stage interactions, to conceptualize new experience-focused design [10,11]. Unlike scenario methods in Model-Based System Engineering (MBSE), which are technically detailed in use case diagrams, sequence diagrams, or activity flows formalized in Systems Modeling Language (SysML)[12,13], trust-aware blueprinting scenarios do not aim to capture structured descriptions of system behavior and system verification. In contrast with medical scenario scripts used in emergency medicine and healthcare education [9], which focus on maintaining clinical accuracy, these scenarios are framed in plain language without high-fidelity medical details.

Fig. 5 illustrates the workflow for application of the Trust-Aware Human-Agent Interaction Blueprinting with stakeholders during the SME workshops, along with the CPH Trust Maps [7,16].

Fig. 5. Workflow with trust-aware scenarios [7].

3.1.1 Defining Context and Personas

The mission context was defined through stakeholder meetings and informed by selections from the NASA STI Program Report [19], which considered multiple mission profiles, including technical constraints and operational aspects. We focused on a particularly challenging scenario: an opposition-class mission with four astronauts, in which only one descends to the surface.

Within this context, four astronaut personas and one agent persona were co-developed during the SME workshop (see Fig. 6). The SMEs were crucial in informing persona development. For the agent persona, the SMEs' contributions captured key insights into the agent's characteristics as perceived from each astronaut's perspective.


Fig. 6. Persona development [7].

3.1.2 Trust-Aware Interaction Blueprinting

To investigate human-agent mutual trust dynamics during the interaction, building on the scenario-based methods, we developed a scenario-based method to map astronaut–agent trust dynamics through trust-aware scenario blueprinting. Fig. 7 illustrates the application of the trust-aware blueprinting worksheet template, designed for the stakeholder activity focused on human-agent trust dynamics. Along with capturing the human-agent interaction, it is designed to capture CPH trust directions: $T_{H\rightarrow A}$ (blue fields) and $T_{A\rightarrow H}$ (yellow fields). The arising issues related to $T_{A\rightarrow A}$ are represented in the bottom row in relation to resources.

Within the activity, diverse stakeholders were divided into small teams. The researcher instructed the SME teams to role-play astronaut–agent medical interactions within the target mission context. Each SME took on a role based on the personas developed in the previous step, acting either as the affected astronaut, the agent, or other crew members who might be involved: Medical Officer (MO) or Point of Care (PoC). The medical challenges for the scenarios were drawn from NASA's IMPACT list of medical conditions most likely to occur during exploration missions[20]. Scenarios were selected to represent a range of conditions and to occur at different mission stages (e.g., transit to Mars, surface EVA, return leg). As the interactions unfolded, the facilitator prompted SMEs to reflect on trust levels throughout the role-play. Trust dynamics were recorded on worksheets, and emerging themes around trust challenges were identified and discussed.

Fig. 7. Trust-aware interaction scenarios [7].

3.1.3 Trust factors and interface implications

Following the capturing of SME role-play scenarios, the next step of the workflow involves consideration of trust factors related to each of the trust directions: $T_{H\rightarrow A}$, $T_{A\rightarrow H}$, and $T_{A\rightarrow A}$. Fig. 8 shows the application of the CPH trust maps as tools designed to systematically consider astronaut agent trust in the trust-driven requirement design and identification of trust-gaps in human-agent systems [7,16]. The CPH trust maps, co-produced with space industry stakeholder interviews [7,16], provide a taxonomy of trust for each of the three trust directions. The trust factors are organized within cognitive-affective and mission-personalized trust dimensions for $T_{H\rightarrow A}$, $T_{A\rightarrow H}$, and related to model performance, training, as well as data processing and collection limitations for $T_{A\rightarrow A}$. (A more detailed derivation of the CPH trust maps is provided in a manuscript currently under submission [16] .)

Fig. 8. Associating trust factors using CPH Trust Maps [7,16].

The CPH trust maps are used to associate trust factors influencing trust dynamics during the interaction from both human and agent perspectives. Considering relevant trust factors within each scenario blueprint allows specification of trust-driven needs and considerations, informing the next stage – trust-aware early concept generation and testing.

3.1.4 Trust-Aware Iterative Interface Prototyping

The trust-aware scenarios provided a backdrop for concept ideation, allowing collaborative early concept testing within the target context and with the developed personas, enabling iterative refinement of the interface. By explicitly considering trust during interactions and addressing trust-driven needs, the design process could focus on design for appropriate reliance and a hands-on approach to 'testing' high-level concepts within these scenarios, providing insights to inform subsequent iterations of the interface.

Fig. 9. Iterative trust-driven interface design conceptualization within the scenarios [7].

3.2 Trust Dynamics in Astronaut-Agent Interaction

To illustrate different aspects of human-agent trust dynamics during the interaction, we present key insights from two scenarios generated through the stakeholder sessions. Unlike traditional scenario-based medical simulation, which focuses on procedural fidelity and clinical accuracy, decision-making under stress, and crisis management [9], the primary role of the participatory HCD role-play scenarios was to highlight trust challenges around human-agent interaction. Thus, in trust-aware scenario mapping, the key focus was not clinical accuracy, but on human-agent collaboration and trust dynamics throughout the interaction. This means that in some instances, the diagnosis has been replaced with 'Condition A and B'. This choice was deliberate to redirect SME focus from clinical and procedural accuracy towards the human-agent trust during the interaction. This allowed us to prioritize the defining trust challenges, associated trust factors, and consequently, define trust-driven opportunities for system/interface design.

3.2.1 Scenario A: Declining Trust and Escalation

The scenario provided insights into how declining agent trust affects interaction architecture, role reallocation, and emergency protocols. During the outbound journey, an astronaut assigned to a Mars surface EVA begins showing signs of severe sleep deprivation, which progressively develops into acute confusion. Concern about medical privacy prevents the astronaut from reporting symptoms early. The agent detects anomalies in sleeping patterns but encounters inconsistencies across data sources. The agent raises concerns about the astronaut's well-being, but the astronaut refuses to engage. As impairment escalates, agent trust declines, and the agent's automation increases. The agent escalates to a medical officer for safety and stabilization and issues an alert to the ground medical team. Table 1 illustrates two example interaction challenges with key associated trust factors considerations, and corresponding interface/system design opportunities.

Table 1. Selected interaction challenges with derived trust-driven opportunities.

Trust	Interaction Challenge	Key CPH Trust Factor Considerations	Trust-Driven Interface/System Design Opportunities
$T_{H\rightarrow A}$ $T_{A\rightarrow A}$	Communicating Observations	FH17: Judgment-Free Consultation FH14: Privacy of Medical Information	*Support privacy* through providing separate Individual and Crew Health Suites. In low-risk situations, the agent could issue an 'excuse note' for the astronaut and support early intervention.
$T_{H\rightarrow A}$ $T_{A\rightarrow H}$ $T_{A\rightarrow A}$	Escalation	FH23: Context-Aware Team Communication FH6: Simulated Mission Training FA1: Mission Risk Profile FA2: Individual Risk Profile	*Pre-Define Levels of Crew Involvement.* In high-risk situations, the agent involves the Crew (MO, PoC) based on the pre-agreed protocol defined with the crew's involvement during the pre-mission crew training activities.

3.2.2 Scenario B: Trust-Adaptative Automation

The scenario captured a complex interaction that takes place across multiple locations and devices (including audio, HUD, tablet, and status preview screen), but also involves various human-agent roles throughout the process and switching of human operators. The astronaut initiates the interaction during the Mars EVA when they experience sharp hip pain. Initially, the interaction takes place over a voice interface and HUD. To account for the astronaut's discomfort, the agent increases the automation, simplifies the interaction and interface, administers a pain reliever, and notifies the rest of the crew. After astronauts return to the habitat, the agent transitions to assisting the Medical Officer (MO) through ultrasound-guided diagnostics. In Table 2, we list selected interaction challenges derived from the scenario, with associated trust factors and design opportunities.

Table 2. Scenario B: Selected interaction challenges with derived trust-driven opportunities.

Trust	Interaction Challenge	Key CPH Trust Factor Considerations	Interface/System Design Opportunities
$T_{A\rightarrow H}$ $T_{H\rightarrow A}$	*Human performance limitation (EVA)*	FH9: Interface Adaptability FA8: Real-Time Biometric Data FA2: Individual Risk Profile	*Dynamic automation and trust-adaptive interface/interaction.* Automation supported by trust-adaptive interfaces and interactions, with system designs that enable increased automation and reduced interaction detail.
$T_{H\rightarrow A}$ $T_{A\rightarrow A}$	*Decision-Support for MO*	FH16: Shared Decision-Making FH16: Agent Confidence Levels MH18: Resource Visualization FS4: Uncertainty Levels	*Illustrate risk visually to support shared decision making.* The agent visually represents the likelihood and consequences of different diagnostic options.

3.3 Trust-Driven Interface Design Recommendations

In this subsection, we synthesize insights from the scenarios and stakeholder workshops into a set of design recommendations for human–agent systems and interfaces. These recommendations highlight the paradigm shifts needed to foster justifiable human–agent trust, while identifying opportunities for role allocation that leverage the complementary strengths of humans and agents to support effective CPH collaboration.

3.3.1 From Standardization to Design for Personalization and Delight

Current crew interface design follows standardization guidelines to minimize training and reduce errors by providing consistency across systems [21]. Exploration medical systems design requires a shift in interface and interaction design towards personalization and customization. Instead of planning for frequent crew rotation and prioritizing standardization of systems, the exploration LDHS missions will need a strong emphasis on designing systems that cater to the specific health needs of the individuals selected for the mission. To prioritize prevention, individual-specific interactions with health systems will be required to minimize risks and maximize performance, given the health impact of the long-duration mission. The frequent engagement with health systems during prevention interactions will require a strong emphasis on designing for adherence and comfort of use—'designing for delight' [14].

Given the strong emphasis on prevention, the agent's primary role will involve monitoring astronauts' health to gather comprehensive biomedical information. Taking on the role of a personal health advisor, trainer, or coach, the agent's relationship with the astronaut aids early detection of potential threats to performance and implementing early interventions for optimizing health and performance. It requires close relationships with individuals and a comprehensive awareness of the mission context, from resources to crew dynamics. It involves both 'passive' observation and routine interaction with the human. This will mean a design of the system that is not only aware of the individual astronaut's health characteristics, but also highly customizable—from the agent's persona and interaction style to the choice between medical or general language and the level of clinical detail—to account for the different needs of astronauts with clinical and non-clinical backgrounds.

3.3.2 From Training People to Fit the Design to Designing Training to Support Customization

By the time of the mission, mutual human-agent trust needs to be already established. There is an opportunity to design systems with a dedicated CPH training that is integrated into the pre-mission training protocols [8]. The design of human-agent interaction within designated training to facilitate building CPH trust can include:

- Individual astronaut health profiling and onboarding, where the astronaut gradually onboards the system, gradually customizes the agent's persona and interaction preferences, while the agent builds an individual health profile based on health history and biomonitoring data;
- Simulated microgravity training, which guides astronaut performance training activities, where the agent gathers information about the physiological response under stress, such as variable-G exposure.

To maximize crew autonomy, the pre-mission training period can also be designed to facilitate CPH team training, including simulated emergency handling to inform mission medical space design. Training activities can also inform future refresher training. Unused skills can degrade over time in humans, but the agent is well-positioned to facilitate refresher training to keep the astronaut's medical training current. Without a human limit on knowledge retention, it can provide on-demand medical training sessions and facilitate team training or medical emergency handling.

3.3.3 From Monitoring and Compliance to Design for Prioritizing Privacy

To support prevention and increase the uptake of early interventions, future systems can leverage privacy prioritization [8]. Concerns about the privacy of medical information, present from early human spaceflight and aviation [22,23], can lead to hesitance in early reporting of medical issues. Stakeholder workshops, SME interviews, and human–agent interaction scenarios highlighted the opportunity to foster $T_{H\to A}$ by designing systems in which the agent can provide judgment-free consultation, making astronauts more inclined to disclose sensitive concerns at an earlier stage [8]. In practice, that could mean separating individual astronaut interface and team-facing interfaces. In practice, this could mean, for example, separating the individual astronaut interface from the team-facing interfaces, as well as pre-defining required Levels of Involvement, determining when the involvement of a Medical Officer or Crew is needed. With privacy prioritization, transparent escalation protocols could be established during the CPH training period before the mission.

3.3.4 From Preventing Trust Breach to Designing for Trust Recovery

Assuming that mishaps are bound to happen, the system design needs to include dedicated trust-recovery interactions. Lack of transparency of what went wrong can quickly erode trust, particularly when the astronaut cannot easily understand the agent's reasoning or limitations. There is an opportunity to design systems for issue explainability, where the agent visualizes the issues it encounters in relation to $T_{A \rightarrow A}$ trust factors (e.g., inconsistency across sources, error in external system, or uncertainty levels). By surfacing issues, the agent helps the crew recalibrate their reliance, re-establish mutual trust, or deploy an issue-specific trust-recovery protocol.

3.3.5 From Medical Guidance to Design for Shared Decision-Making and Dynamic Automation

The diversity and complexity of comprehensive care for astronauts on LDFS necessitate consideration of multiple interdependent tasks across the medical decision-making process, from initial symptom observation to treatment delivery. Likelihood and uncertainty around diagnostic and treatment options need to be weighed alongside optimizing the allocation of scarce resources over an extended mission duration. With an astronaut-physician likely present on board, one of the key roles of an agent will entail providing highly specialized assistance to the human doctor [24]. In cases where a physician or medical officer is absent or incapacitated, however, the agent will need to assume greater responsibility and automation. Moreover, roles may shift dynamically during an interaction—for instance, when additional crew join the response. Accordingly, interfaces and interactions must be designed to support dynamic automation and adaptive role allocation throughout the decision-making process.

3.3.6 From Uniform Interaction to Design for Variety

The diversity of medical tasks in long-duration missions necessitates a corresponding diversity in information presentation and interaction style. To optimize collaboration, interfaces must be appropriately catered to the context of use and the task at hand. Some interactions may benefit from a doctor–patient style dialogue, while others require comparative overviews or a structured table [8]. Variety also plays a critical role in managing cognitive load: In non-time-critical contexts, more detailed information and options may be appropriate, whereas in urgent or high-stress situations, the interface should present only the most essential information at the right time. Designing for variety enables flexible, context-sensitive interaction styles that align with task demands and crew needs.

3.3.7 From Designing for Operator's Trust to Designing Trust-Adaptive Interaction

Design for trust-adaptive interaction shifts the focus from simply supporting the operator's trust to dynamically adjusting the agent's behavior based on real-time trust assessments [7,8]. Assistance is tailored to the human's current state, environment, and prior knowledge of the astronaut, allowing the agent to modulate information and guidance according to the operator's needs. In high-trust situations, collaboration is rich and detailed, whereas in low-trust or time-critical scenarios, the agent provides more direct instructions and simplifies the interface. This dynamic adaptation supports mutual trust calibration while allowing the human to override or adjust the level of support as needed.

4. Conclusions

This paper emphasized mutual trust as a first-class component in astronaut–agent medical collaboration. We introduced the trust-aware scenario blueprinting workflow and demonstrated its application in facilitating transdisciplinary dialogue. By treating trust as a lens for examining astronaut–agent interaction, we presented a way to derive interaction, interface, and system design needs that support appropriate reliance and effective teamwork. Trust-aware scenarios, used as boundary objects, enabled diverse stakeholders to jointly investigate trust dynamics and surface design challenges and opportunities that might otherwise remain implicit.

We presented transdisciplinary recommendations for designing CPH-oriented interfaces that leverage the complementary strengths of humans and agents while embedding mutual justifiable trust as a core consideration for future medical systems.

Our ongoing work includes the design of trust-driven and trust-adaptive CPH Clinical Decision Support (CDS) interfaces, as well as the development of CPH trust diagnostics methods for existing systems. Our future work will explore the neuroergonomics of dynamic interfaces for astronaut–agent shared decision-making.

The design of future CPH-oriented exploration medical systems requires diverse stakeholder collaboration, and integration of human-centered design expertise to facilitate the transdisciplinary dialogue to ensure systems support trust-building and enable effective human–agent medical teaming.

Acknowledgements

This research was supported by the NASA Langley Research Center, the Goddard Space Flight Center, and the Jet Propulsion Laboratory, California Institute of Technology, through a grant awarded to the first and third authors (Grant Award Number: 22-1-T10.05-2368LaRC), and the Aerospace Medical Association Foundation Jeffrey Myers Young Investigator Award presented at the Annual Scientific Meeting of the Aerospace Medical Association to the first author. The funding organizations did not participate in the writing of this publication, nor in the collection, analysis, or interpretation of the data. The content, findings, and conclusions presented are the authors' own and should not be interpreted as the official position of NASA and AsMA.

The authors thank the SMEs and stakeholders for their generous contributions of time and expertise during the interviews, design meetings, and SME workshops.

CRediT author statement

Anna B. Wojdecka: Conceptualization, Funding acquisition, Methodology, Investigation, Data Curation, Formal Analysis, Visualization, Writing–Original Draft, Writing–Review & Editing. Tibor Balint: Validation, Writing–Review & Editing. Don Platt: Supervision, Funding acquisition, Writing–Review & Editing.

Disclosure of Interest. The authors declare that they have no competing interests.

References

[1] D. Hamilton, K. Smart, S. Melton, J.D. Polk, K. Johnson-Throop, Autonomous Medical Care for Exploration Class Space Missions, J. Trauma: Inj., Infect., Crit. Care 64 (2008) S354–S363. https://doi.org/10.1097/ta.0b013e31816c005d.

[2] S.L. Johnston, K.T. Smart, J.M. Pattarini, Principles of Clinical Medicine for Space Flight, (2020) 327–353. https://doi.org/10.1007/978-1-4939-9889-0_10.

[3] T.A. Taddeo, S. Gilmore, C.W. Armstrong, Principles of Clinical Medicine for Space Flight, in: Springer, New York, 2019: pp. 201–231. https://doi.org/10.1007/978-1-4939-9889-0_6.

[4] NASA ExMC, Human Research Roadmap (HRR) Gaps. Medical-501: We need to develop integrated exploration medical system models for the Moon and Mars, (2024). https://humanresearchroadmap.nasa.gov/gaps/gap.aspx?i=714#.

[5] N. Kraetschmer, N. Sharpe, S. Urowitz, R.B. Deber, How does trust affect patient preferences for participation in decision-making?, Heal. Expect. 7 (2004) 317–326. https://doi.org/10.1111/j.1369-7625.2004.00296.x.

[6] R.C. Mayer, J.H. Davis, F.D. Schoorman, An Integrative Model of Organizational Trust, Acad Management Rev 20 (1995) 709. https://doi.org/10.2307/258792.

[7] A. Wojdecka, D. Platt, Transdisciplinary Design of Trust-Driven Clinical Decision Support Interface for Cyber-Physical Human Teams in Long-Duration Human Spaceflight, in: IAA SCITECH 2025 Forum, 2025. https://doi.org/10.2514/6.2025-2094.

[8] A.B. Wojdecka, T. Balint, D. Platt, Towards Trust-Driven Trust-Adaptive Astronaut-Agent Medical Collaboration Interfaces, in: HCI International 2024 – Late Breaking Papers, Gothenburg, Sweden, 2025.

[9] D.M. Gaba, S.K. Howard, K.J. Fish, B.E. Smith, Y.A. Sowb, Simulation-Based Training in Anesthesia Crisis Resource Management (ACRM): A Decade of Experience, Simul. Gaming 32 (2001) 175–193. https://doi.org/10.1177/104687810103200206.

[10] M.J. Bitner, A.L. Ostrom, F.N. Morgan, Service Blueprinting: A Practical Technique for Service Innovation, Calif. Manag. Rev. 50 (2008) 66–94. https://doi.org/10.2307/41166446.

[11] G.L. Shostack, Designing Services That Deliver, Harvard Business Review 62 (1984) 133–139. https://hbr.org/1984/01/designing-services-that-deliver.

[12] I. Jacobson, Object-oriented software engineering, Association for Computing Machinery, New York, NY, United States, 1991.

[13] S. Friedenthal, A. Moore, R. Steiner, A Practical Guide to SysML, Third Edition: The Systems Modeling Language, Morgan Kaufmann Publishers Inc., San Francisco, CA, United States, 2014.

[14] R. Glanville, The purpose of second-order cybernetics, Kybernetes 33 (2004) 1379–1386. https://doi.org/10.1108/03684920410556016.

[15] A.B. Wojdecka, A. Hall, T. Balint, Designing Out Unwanted Healthcare Futures: A New Framework For Healthcare Design Innovation With Intent, Ds 117 Proc 24th Int Conf Eng Prod Des Educ E Pde 2022 Lond South Bank Univ Lond Uk 8th - 9th Sept 2022 (2022). https://doi.org/10.35199/epde.2022.110.

[16] A. Wojdecka, O. Doule, D. Platt, N. Alexandrov, Trust by Design: A Taxonomy and Tool for Astronaut–Agent Systems, Manuscript in Preparation (n.d.).

[17] T.S. Balint, P. Pangaro, Design space for space design: Dialogs through boundary objects at the intersections of art, design, science, and engineering, Acta Astronaut. 134 (2017) 41–53. https://doi.org/10.1016/j.actaastro.2017.01.029.

[18] K.T. Simsarian, Take it to the next stage, CHI '03 Ext. Abstr. Hum. Factors Comput. Syst. - CHI ’03 (2003) 1012–1013. https://doi.org/10.1145/765891.766123.

[19] Valinia, A., Allen, J. R., Francisco, D. R., Minow, J. I., Pellish, J. A., & Vera, A.H. (2022) (No. NESC-RP-20-01589). Safe Human Expeditions Beyond Low Earth Orbit (LEO), No. NESC-RP-20-01589 (2022).

[20] A.J. Kreykes, R. Suresh, D. Levin, D.C. Hilmers, Selecting Medical Conditions Relevant to Exploration Spaceflight to Create the IMPACT 1.0 Medical Condition List, Aerosp. Med. Hum.

Perform. 94 (2023) 550–557. https://doi.org/10.3357/amhp.6199.2023.

[21] NASA, NASA-STD-3001, Space Flight Human-System Standard Volume 1: Crew Health, (2025). https://standards.nasa.gov/standard/NASA/NASA-STD-3001-VOL-1.

[22] C.A. Berry, G.W. Hoffler, C.A. Jernigan, J.P. Kerwin, S.R. Mohler, History of Space Medicine: The Formative Years at NASA, Aviat., Space, Environ. Med. 80 (2009) 345–352. https://doi.org/10.3357/asem.2463.2009.

[23] T.-E. Strand, N. Lystrup, M. Martinussen, Under-Reporting of Self-Reported Medical Conditions in Aviation: A Cross-Sectional Survey, Aerosp. Med. Hum. Perform. 93 (2022) 376–383. https://doi.org/10.3357/amhp.5823.2022.

[24] R.T. Sutton, D. Pincock, D.C. Baumgart, D.C. Sadowski, R.N. Fedorak, K.I. Kroeker, An overview of clinical decision support systems: benefits, risks, and strategies for success, Npj Digit. Med. 3 (2020) 17. https://doi.org/10.1038/s41746-020-0221-y.