Narrative Review Article

A Smarter Path to Mars: A Conceptual Framework for Human-AI Teamwork in Surface Exploration

Junyao Zhang[1] and Hongsheng Shang[2]

[1]*Bonita High School, 3102 D St, La Verne, CA 91750, United States;*
[2]*Dexter Southfield School, 20 Newton St, Brookline, MA 02445, United States*

ABSTRACT

Future human missions to Mars will place astronauts in a world that is scientifically rich but physically unforgiving. The Martian surface has a thin atmosphere, extreme temperature swings, dust activity, radiation exposure, delayed communication with Earth, limited resupply, and full dependence on engineered life-support systems. These conditions make Mars exploration a safety-critical, distributed teamwork problem rather than a simple task-planning problem. This article develops a conceptual framework for human-artificial intelligence (AI) teamwork in Mars surface exploration. No detailed mathematical model is proposed, no AI system is trained, no operational performance is claimed, and no crewed test has been performed. Instead, the contribution is a research-grounded engineering concept organized around Figure 1, in which mission readiness, in-mission support, human review, spacecraft and habitat architecture, and data-to-value feedback form a closed safety loop. The framework argues that AI should not replace astronauts or mission-control teams; rather, AI should act as a safety translator that turns environmental data, system telemetry, robot reports, and science priorities into explainable options for human approval. The proposed concept adds engineering soundness by mapping Mars hazards to decision-support functions, defining human-authority requirements, identifying safety and planetary-protection guardrails, and laying out a staged validation pathway from concept review to tabletop exercises, analog missions, digital twins, and eventually certified operational systems. The central message is that the smartest path to Mars is neither full automation nor unaided human courage, but disciplined human-AI-robot teamwork that helps explorers remain safe, aware, ethical, and scientifically productive.

Keywords: Mars exploration; human-artificial intelligence teaming; autonomous systems; safety-critical operations; planetary protection

INTRODUCTION

Human exploration of Mars is one of the most ambitious goals in space exploration. The National Aeronautics and Space Administration's (NASA's) Moon to Mars Architecture describes future exploration as an objectives-based systems-engineering effort for long-term human and robotic exploration beyond Earth,

Corresponding author: Junyao Zhang, E-mail: Cindyzhjy@gmail.com.

Accepted July 14, 2026
https://doi.org/10.70251/HYJR2348.44269282

including integrated human and robotic methods for science, safe operations, and eventual Mars infrastructure (1). A human mission to Mars would not be only a transportation challenge. It would also be a continuous decision-making challenge in which astronauts, robots, habitats, spacesuits, science instruments, and Earth-based mission control must operate as one distributed team.

Mars is dangerous precisely because it combines many risks at once. The planet is cold, dusty, dry, low-pressure, and far away. It does not provide breathable air, easy shelter, rapid rescue, or real-time communication with Earth. Astronauts would work inside a web of engineered protections: spacesuits, habitats, life-support systems, rovers, power systems, sensors, and emergency procedures. If one part of that web changes, the meaning of many other parts can change as well. A dust event may affect power and visibility; a communication gap may slow expert support; a sensor warning may interact with crew fatigue or rover availability. Safe exploration therefore depends on maintaining situation awareness under uncertainty.

Artificial intelligence (AI) can help with this challenge, but only if its role is defined carefully. AI can monitor many data streams, organize alerts, detect conflicts, summarize risk, suggest possible responses, and preserve a record of why a decision was made. Humans bring scientific curiosity, ethical responsibility, risk judgment, common sense, and the ability to recognize context that may not appear in sensor data. Robots and habitat systems can gather information or perform work that would otherwise expose astronauts to unnecessary danger. The proposed framework links these strengths through a human-centered autonomy loop rather than through an automation-first design.

The central research question is: How can a human-AI collaboration loop support safer and more organized Mars surface exploration while preserving human authority, scientific purpose, and planetary protection? The answer is developed as a conceptual framework rather than as a detailed engineering model. The manuscript does not claim that a particular algorithm is ready for Mars operations. It instead proposes a structured way to think about roles, responsibilities, guardrails, and future validation.

The contribution is fourfold. First, the framework explains why Mars surface exploration creates a need for local AI decision support. Second, it describes why humans must remain central despite the usefulness of autonomy. Third, it translates Figure 1 into a generalized concept of operations for astronauts, AI assistants, robots, habitats, and mission control. Fourth, it proposes engineering design principles and an assurance pathway that could guide later research projects, expert review, analog testing, and formal simulation. The innovation is not a new algorithm, but a clearer responsibility architecture for human-AI teamwork in an extreme environment.

The framework was developed through conceptual synthesis. Sources were selected from three areas: (1) NASA and Jet Propulsion Laboratory (JPL) documents on Mars exploration, human spaceflight hazards, life-support systems, mission planning, planetary protection, and AI use; (2) Mars rover autonomy and planning literature; and (3) human factors and systems-engineering literature on situation awareness, trust, automation, and safety assurance. This approach is appropriate for an early-stage concept paper because the goal is to organize existing knowledge into a useful design framework, not to report experimental results.

The scope is intentionally bounded. No astronaut crew has tested the framework. No operational AI assistant has been built. No detailed mission timeline, resource model, machine-learning architecture, or surface logistics optimizer is presented. No numerical improvement in safety, efficiency, or science return is claimed. The concept should therefore be read as a pre-design framework: a set of structured questions, roles, requirements, and validation steps that could guide later engineering work.

A concept can still be scientific and engineering sound when it states its assumptions, links its logic to the literature, identifies failure modes, and describes what evidence would be needed before operational use. NASA's Systems Engineering Handbook emphasizes requirements, logical decomposition, verification, validation, risk management, interface management, and decision analysis as part of disciplined system development (2). Those ideas are reflected here by converting a simple collaboration loop into roles, information flows, requirements, guardrails, and future tests.

Table 1 clarifies the boundary between what the framework includes and what it does not claim. This distinction is important for peer review because a concept framework can be useful only when it avoids implying that an untested assistant is ready for flight or that performance improvements have already been demonstrated.

Table 1. *Scope boundaries for the conceptual framework. The table distinguishes the early-stage human-artificial intelligence (AI)-robot teamwork concept proposed in this paper from claims that are not made, such as flight-ready software, a finished mission architecture, or validated performance data. Abbreviations: AI, artificial intelligence; JPL, Jet Propulsion Laboratory; NASA, National Aeronautics and Space Administration.*

Element	Included in the framework	Not claimed
Concept type	Generalized human-AI-robot teamwork framework for Mars surface exploration.	A finished mission architecture or flight-ready software design.
Technical depth	Role definitions, information flows, design principles, safety guardrails, and validation pathway.	Detailed algorithms, trained models, code, optimization results, or hardware specifications.
Evidence basis	Literature-grounded reasoning from NASA/JPL sources, rover autonomy, human factors, and systems engineering.	Crew test results, analog mission performance data, or operational reliability statistics.
Primary value	A structured concept that helps researchers and mission designers discuss responsible AI in Mars exploration.	Proof that AI will improve a specific Mars mission metric.

WHY HUMANS NEED AI PARTNERS ON MARS

Mars is dangerous for unaided human activity

Mars is not an environment where humans can walk, breathe, or work without engineered protection. NASA Science describes Mars as a dusty, cold desert world with a very thin atmosphere. Its atmosphere is mostly carbon dioxide, nitrogen, and argon, and it provides little protection from incoming objects. Temperatures can rise to about 20 degrees Celsius or fall to about -153 degrees Celsius, and winds can create dust storms that cover much of the planet (3). These conditions make every surface activity dependent on suits, habitats, rovers, environmental sensors, communication systems, and emergency plans.

Human spaceflight also involves broader hazards. NASA Human Research Program identifies five major hazard categories for deep-space missions: space radiation, isolation and confinement, distance from Earth, gravity fields, and hostile or closed environments (4). These hazards can interact rather than occur separately. For example, a dust storm could reduce power margins, limit visibility, delay surface travel, and increase crew stress at the same time. A life-support anomaly could become urgent because the habitat is the crew's only safe internal environment. AI support is valuable not because it removes danger, but because it can help the crew see how risks combine.

Life support is especially important. NASA describes the Environmental Control and Life Support System (ECLSS) as providing or controlling atmospheric pressure, fire detection and suppression, oxygen levels, ventilation, waste management, and water supply (5). A Mars habitat would connect these functions to power, maintenance, crew health, spare parts, and surface operations. A human crew may not be able to watch every subsystem continuously while also planning exploration and science. AI-assisted monitoring can act as an attention amplifier by filtering routine data, highlighting weak signals, and helping the crew focus on the most safety-relevant changes.

Distance from Earth makes local decision support necessary

Mars is too far away for real-time remote control. The European Space Agency (ESA) explains that one-way light-time between Earth and Mars can vary from about 4.3 minutes to about 21 minutes depending on planetary positions (6). NASA similarly notes that a crew traveling to Mars would face communication delays of up to about 20 minutes one way, equipment failures, medical emergencies, limited supplies, and a need for self-sufficiency with minimal support from Earth (7). NASA technical work has also identified communication delays, disruptions, and blackouts as important challenges for crewed Mars missions (8).

These delays do not make Earth-based mission control unimportant. Mission control would remain a major source of expertise, long-term planning, analysis, and independent review. However, a Mars crew may need to respond before a complete question-and-answer cycle with Earth is possible. Local decision support could help astronauts answer urgent questions: What has changed since the last plan? Which safety limits are near violation? Which robot can investigate before a person

goes outside? Which information should be sent to Earth for later review? Delayed communication therefore strengthens the case for local human-AI teamwork.

AI and autonomy already support space operations

AI and autonomy are already part of space operations, although current systems are not equivalent to a human-rated Mars surface assistant. NASA's 2026 AI Inventory states that NASA is using AI in research, tools, models, and public-benefit projects, with many activities still in formulation or development (9). NASA's AI use-case review describes applications across autonomous exploration and navigation, mission planning and management, environmental monitoring, data management, and Mars rover operations (10). NASA Ames Research Center also describes planning and scheduling research in areas such as AI planning, combinatorial optimization, constraint satisfaction, and multi-agent coordination (11).

Rover examples show both the promise and the limits of autonomy. Mars Exploration Rover activity planning required plans that satisfied resource, time, and safety constraints (12). The AEGIS system allowed Mars rovers to select science targets autonomously, increasing science opportunities when the ground team was not immediately available (13,14). In 2026, JPL reported that Perseverance completed drives on Mars planned by AI-generated waypoints, after commands were verified using a digital twin and extensive telemetry checks (15). These examples support the idea that intelligent systems can help off-world missions, but they also show that validation, constraint checking, and human oversight remain essential.

A future human surface assistant would require even stronger safeguards than robotic science autonomy because its recommendations could affect crew safety. It should therefore be treated as a safety-critical decision-support system. The key question is not whether AI can produce a recommendation, but whether humans can understand, challenge, verify, and responsibly use that recommendation under Mars conditions.

Humans must remain central

The danger of Mars does not imply that AI should replace astronauts. In many ways, danger makes human authority more important. Humans notice context that may be difficult to encode in data: fatigue, stress, unusual sounds, visual impressions, suit discomfort, team morale, scientific curiosity, and ethical obligations. Human factors research emphasizes situation awareness as a foundation for dynamic decision-making, including perception of relevant elements, comprehension of their meaning, and projection of future status (16). AI should strengthen situation awareness rather than narrow it to a single automated answer.

Human-automation research also warns that automation can change how people think. Parasuraman *et al*. (17) described different levels and types of automation, showing that automation can support information acquisition, information analysis, decision selection, or action implementation. Lee and See (18) argued that trust in automation must be calibrated so that people neither underuse useful automation nor over-rely on flawed automation. For Mars, this means the AI assistant should be designed to invite questions, reveal uncertainty, show alternatives, and support human judgment. A confident-looking but unexplained recommendation would be dangerous if it encouraged astronauts to stop thinking critically.

The best role for AI in the proposed framework is therefore partner, not commander. AI should gather and translate information; humans should decide what risks are acceptable, what science matters most, and when a recommendation must be overruled. This division of responsibility is not anti-technology. It is a practical design choice for a mission environment where safety, ethics, science, and human survival are inseparable.

Conceptual gap and innovation

Many discussions of AI in space focus on either narrow optimization or high-level autonomy. A scheduling tool might ask how to complete tasks faster; a rover autonomy tool might ask how to navigate without waiting for Earth; an AI science tool might ask how to select targets. Those are valuable questions, but future human Mars exploration also needs a broader framework for responsibility. The proposed concept treats AI as a safety translator between the Martian environment and human decision-makers.

Table 2 summarizes the research basis behind the framework and connects each major source area to an engineering implication. The table shows that the proposal is not based on a single technology trend; it combines Mars environmental hazards, communication limits, life-support dependence, rover autonomy, human factors, systems engineering, and planetary-protection requirements.

A safety translator does more than compute. It turns danger and data into understandable choices. It separates facts from estimates, marks uncertainty, identifies

Table 2. *Research basis used to support the generalized concept framework. The table summarizes the major source areas that informed the concept and explains how each source area translates into a design implication for Mars surface exploration. Abbreviations: AI, artificial intelligence; ECLSS, Environmental Control and Life Support System; ESA, European Space Agency; JPL, Jet Propulsion Laboratory; NASA, National Aeronautics and Space Administration.*

Research basis	What current sources show	Implication for the concept
Mars environment	Mars has a thin atmosphere, extreme temperatures, dust, and limited natural protection (3).	Humans need habitats, suits, sensors, robots, and decision support rather than unaided activity.
Human spaceflight hazards	Radiation, isolation, distance, gravity changes, and closed or hostile environments are major risks (4).	AI support should focus on safety awareness, not only efficiency.
Communication delay	Earth-Mars communication can involve many minutes of one-way delay and possible disruptions (6, 8).	Crews need local tools that help them reason before Earth can respond.
Planning and autonomy	Rover planning and AI-assisted Mars operations show the value of constraint-aware autonomy and validation (12, 15).	A Mars assistant should provide recommendations, explanations, and verification support.
Life support	ECLSS controls pressure, oxygen, ventilation, water, waste, and fire-related functions (5).	AI should monitor critical systems and treat safety limits as hard constraints.
Human factors	Situation awareness and calibrated trust are central to safe human-automation interaction (16, 18).	AI should make uncertainty visible and preserve human review.
Planetary protection	Planetary protection aims to prevent forward contamination and backward contamination (19).	AI recommendations should include keep-out zones, cleanliness rules, and decision records.

missing information, flags conflicts, records assumptions, and explains which safety or planetary-protection rules matter. The innovation is a shared safety loop: machines provide endurance, speed, and continuous monitoring, while humans preserve judgment, meaning, ethics, and authority. This idea is clear enough for broad audiences to understand and rigorous enough to guide later engineering analysis.

CONCEPTUAL FRAMEWORK

Figure 1 is the organizing concept. It shows human-AI teamwork as a loop rather than a one-time command. The loop begins with mission readiness, continues through in-mission AI support and human choice, connects to spacecraft and habitat architecture, and turns data into value for future decisions. The arrows matter because Mars operations will be dynamic. Every decision changes the next situation; every result produces new evidence; every warning or near miss should improve the next cycle.

The five parts of the loop can be interpreted as five engineering questions. Are the mission, systems, crew, and robots ready? What support can AI provide during the mission? What should humans decide after reviewing the evidence? How do spacecraft, habitats, rovers, and tools constrain the available actions? How does the team turn raw data into useful knowledge for the next decision? These questions make Figure 1 more than a diagram. They turn it into a repeatable operating habit for responsible exploration.

Mission readiness

Mission readiness means that the team begins before an astronaut steps onto the surface. Readiness includes

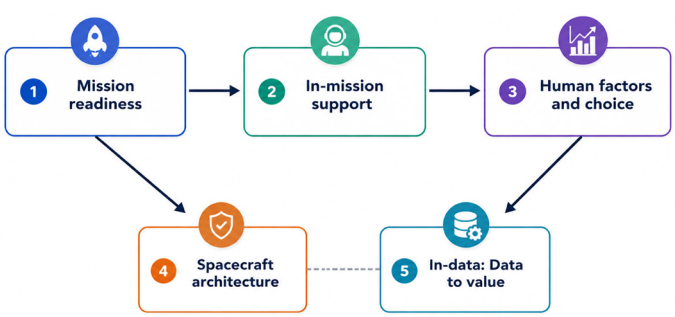

Figure 1. *Conceptual human-AI teamwork loop for Mars exploration. The loop links mission readiness, in-mission support, human factors and choice, spacecraft architecture, and data-to-value feedback.*

hardware status, crew preparation, science priorities, environmental conditions, communication windows, spare parts, emergency plans, and planetary-protection constraints. AI can support readiness by checking whether all required information is available, whether safety margins are acceptable, and whether the proposed activity conflicts with other mission constraints. A useful assistant should also identify missing information rather than hiding uncertainty.

In-mission support

In-mission support is the real-time or near-real-time assistance provided while astronauts, robots, and habitat systems are operating. The AI assistant can combine suit telemetry, rover status, power trends, weather observations, terrain information, habitat readings, and mission objectives into a concise risk picture. Its output should be understandable: not simply an alarm, but a short explanation of what changed, why it matters, what choices exist, and what stop conditions should be watched.

Human factors and choice

Human choice is the control point of the framework. Astronauts and mission leaders should be able to approve, revise, delay, or reject any AI recommendation. The review step allows humans to include context that may not be captured in data, such as crew fatigue, stress, visual impressions, sound, smell, discomfort, team confidence, or scientific judgment. Preserving choice also protects accountability. If a decision involves risk to crew safety, scientific integrity, or planetary protection, the responsible decision-maker must remain human.

Spacecraft, habitat, and robotic architecture

The loop is not only a software concept. It depends on physical architecture: habitats, suits, rovers, robotic arms, sensors, sample tools, power systems, communication relays, and emergency procedures. AI recommendations should respect what the architecture can actually do. For example, if a rover is unavailable, if a habitat sensor is uncertain, if a suit battery is near a limit, or if a sample site requires planetary-protection care, the assistant should adjust its options accordingly. This turns AI support into systems engineering rather than isolated advice.

Data-to-value feedback

Mars missions will produce large amounts of data: images, telemetry, environmental readings, crew notes, maintenance records, rover reports, and science observations. Data become valuable only when they help the team understand the situation and make better decisions. The proposed AI assistant should turn data into value by organizing evidence, preserving decision records, summarizing lessons learned, and helping mission control review decisions after a communication delay. Over time, the loop can build operational memory: a record of what was tried, what worked, what nearly failed, and what should change next time.

CONCEPT OF OPERATIONS FOR HUMAN-AI-ROBOT COLLABORATION

The concept around Figure 1 can be translated into a general concept of operations without creating a detailed mathematical model. The operating idea is that each Mars surface activity begins with a shared situation picture, moves through AI-supported reasoning, requires human review, assigns action to astronauts, robots, habitat systems, or mission control, and records feedback. The aim is to reduce confusion and unnecessary exposure to danger, not to make Mars safe or to prove a performance improvement.

A typical pre-extravehicular activity (pre-EVA) cycle might begin with the AI assistant checking dust conditions, temperature, suit status, rover readiness, power margins, crew workload, planned route, communication windows, and science priority. The assistant could then present a risk summary and several options: proceed as planned, shorten the EVA, send a robot first, delay until better conditions, or change the science target. The human crew would review the evidence, challenge assumptions, and select a course of action. After the activity, the record would include what was decided, why it was decided, what happened, and what should be updated.

The same loop could support habitat maintenance. If a sensor reports an unusual reading, the assistant should not simply generate an alarm. It should compare the reading with recent maintenance, power status, ventilation history, crew observations, and life-support limits. It should ask whether a safe remote check is possible before a human inspection. If conflicting data exist, the assistant should flag the conflict and recommend conservative action until the crew can verify the cause.

Mission control remains part of the team. When Earth cannot respond immediately, the crew can use the loop locally and send a structured decision package to

Earth: situation summary, AI recommendation, human reasoning, chosen action, rejected alternatives, and observed outcome. Mission control can later review the record, improve procedures, and update future decision rules. In this way, communication delay becomes less of a barrier and more of a slower layer of reflection and learning.

Table 3 translates the Figure 1 loop into an operational structure by identifying the primary inputs, AI assistance, human decision point, and engineering safeguard for each stage. Table 4 then grounds the same logic in representative Mars situations, showing how the collaboration loop could support planning and response without claiming a tested mission procedure.

The candidate situations in Table 4 illustrate why the framework must remain flexible. Pre-EVA planning, habitat anomalies, science traverses, dust events, and communication loss all require different information, but the same basic pattern remains: AI organizes evidence, humans accept responsibility for the decision, and robots or systems reduce unnecessary human exposure when possible.

Table 3. *Generalized concept of operations around Figure 1. The table maps each stage of the proposed human-AI teamwork loop to its expected inputs, possible AI assistance, human decision point, and engineering safeguard. Abbreviation: AI, artificial intelligence.*

Loop stage	Primary inputs	AI assistance	Human decision	Engineer safeguard
1. Mission readiness	Mission goals, crew status, habitat readiness, robot status, environmental forecast, communication windows.	Checks completeness, highlights missing data, identifies known constraints and open risks.	Decide whether the planned activity is ready to enter execution review.	Readiness checklist, minimum data rules, and required human sign-off.
2. In-mission support	Suit, rover, power, terrain, weather, science, life-support, and procedure data.	Summarizes risk, identifies conflicts, explains options, and marks uncertainty.	Approve, revise, delay, or reject the suggested action.	Hard safety limits, stop rules, confidence flags, and alternative procedures.
3. Human factors and choice	Crew observations, fatigue, stress, science judgment, ethics, and mission priorities.	Shows evidence and records questions, changes, and rejected options.	Accept risk only when the crew understands the basis and consequences.	Human authority, override ability, and accountability record.
4. Spacecraft architecture	Capabilities and limits of suits, rovers, tools, habitats, sensors, and communication systems.	Checks whether the approved plan matches physical system limits.	Assign action to astronaut, robot, habitat system, or mission control.	Interface checks, configuration control, and resource constraints.
5. Data to value	Results, near misses, telemetry, images, crew notes, and mission-control feedback.	Converts outcomes into lessons for the next cycle and preserves traceability.	Decide whether procedures, thresholds, or priorities need updating.	Decision log, anomaly review, and learning loop.

Table 4. *Candidate Mars situations for applying the collaboration loop. The table gives example use cases in which AI could organize information and present options while astronauts or mission leaders retain authority over safety-relevant decisions. Abbreviations: AI, artificial intelligence; EVA, extravehicular activity.*

Situation	AI support	Human authority	Why collaboration
Pre-EVA planning	Summarize dust, terrain, suit, rover, power, crew workload, route, science goals, and communication-window information.	Approve the EVA, change its goals, shorten it, delay it, or send a robot first.	Reduces avoidable human exposure before astronauts leave the habitat.

Continued Table 4. *Candidate Mars situations for applying the collaboration loop. The table gives example use cases in which AI could organize information and present options while astronauts or mission leaders retain authority over safety-relevant decisions. Abbreviations: AI, artificial intelligence; EVA, extravehicular activity.*

Situation	AI support	Human authority	Why collaboration
Habitat anomaly	Compare sensor readings, recent maintenance, power status, ventilation, spare parts, and life-support limits.	Choose the safest inspection or repair sequence and decide when to request Earth review.	Protects life-support systems while keeping people responsible for risk acceptance.
Science traverse	Highlight route hazards, robot readiness, sample priorities, planetary-protection constraints, and time margins.	Select sites and decide which science opportunities justify exposure to risk.	Balances curiosity, safety, and site protection.
Dust or power event	Detect changing power margins, suggest conservative fallback options, and identify nonessential activities to pause.	Reduce activity, change priorities, enter contingency mode, or wait for Earth guidance.	Supports local response when communication delay prevents immediate Earth control.
Medical or crew-condition concern	Organize available medical data, workload history, communication opportunities, and emergency options.	Decide whether to stop an activity, return to habitat, request medical guidance, or activate contingency procedures.	Keeps human welfare central while helping the crew avoid information overload.

ENGINEERING DESIGN PRINCIPLES AND REQUIREMENTS

A concept paper becomes more engineering sound when it converts vision into requirements. The proposed assistant should be designed for safety before speed. It should not simply optimize for task completion, distance traveled, or science return. Instead, it should help the crew maintain awareness, protect life-support margins, preserve human authority, respect planetary-protection rules, and create a traceable record of decisions.

NASA's ethical AI material emphasizes principles such as fairness, human-centered use, explainability and transparency, accountability, safety and security, and scientific and technical robustness (20). Those principles match the needs of a Mars surface assistant. A system used near human explorers must be explainable enough for astronauts to question it, accountable enough for mission control to audit it, and robust enough to fail safely when information is incomplete.

The design should also avoid two opposite mistakes: under-automation and over-automation. Under-automation would force astronauts to manually monitor too many systems under stress. Over-automation would encourage the crew to accept recommendations without understanding them. The proposed middle path is bounded autonomy: AI may monitor, analyze, summarize, rank, and warn, but human leaders retain decision authority for actions that affect crew safety, scientific priorities, or planetary protection.

A practical interface should therefore present each recommendation as a short safety card: the proposed action, the evidence used, the data freshness, the key uncertainty, the hard limits that must not be violated, the safer alternatives, and the record that should be saved for later review. This keeps the assistant useful under stress while making the reasoning visible enough for astronauts and mission control to challenge.

Figure 2 develops this safety-card idea into a clearer engineering role for AI. Mars danger and mission data are treated as inputs, the AI assistant translates those inputs into explainable options, and human authority remains the approval point. The lower part of the figure emphasizes that robots, habitats, and mission-control teams execute or review approved actions while safety constraints, traceable records, and planetary protection remain active.

Table 5 converts the concept into candidate engineering requirements and guardrails for a future decision-support assistant. These items should be treated as design conditions rather than optional features: without human authority, explainability, safety constraints, uncertainty visibility, and traceable records, an AI assistant could increase risk instead of reducing it.

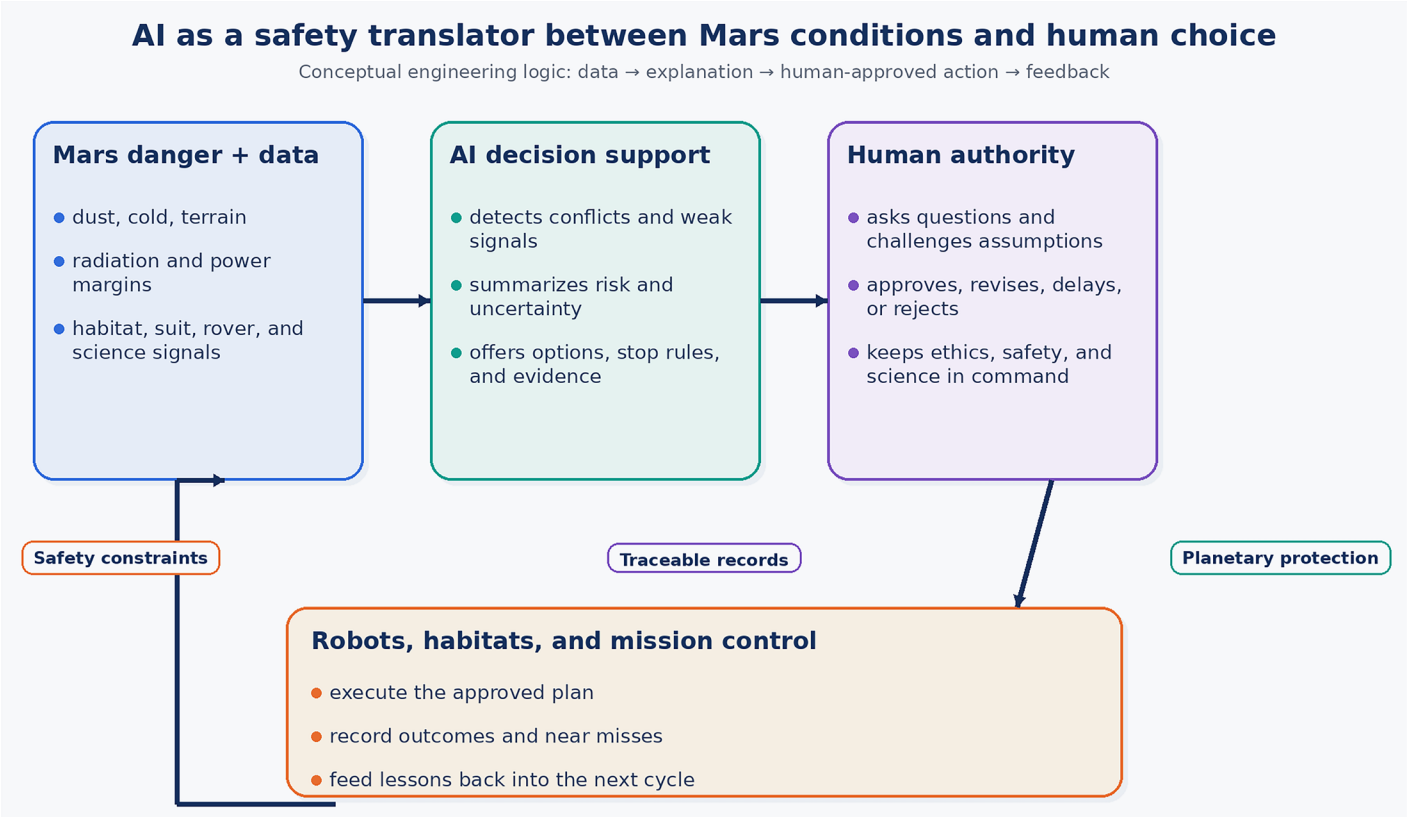

Figure 2. *AI as a safety translator. The assistant converts Mars danger and mission data into explainable options, while humans preserve authority and robots, habitats, and mission control execute or review approved actions.*

Table 5. *Engineering requirements and guardrails for a future Mars AI decision-support assistant. The table lists candidate design requirements that would help ensure an AI assistant remains explainable, bounded, traceable, and subordinate to human authority. Abbreviation: AI, artificial intelligence.*

Requirement / guardrail	Engineering meaning	Reason for Mars relevance
Human authority	Every safety-relevant recommendation must be approvable, revisable, rejectable, and overridable by authorized humans.	Crew safety and risk acceptance are human responsibilities.
Explainable output	The assistant must show evidence used, assumptions made, uncertainty, alternatives, and reasons for a recommendation.	Delayed communication and high stress require quick but understandable decisions.
Hard safety constraints	Life-support, suit, radiation, terrain, pressure, power, fire, and emergency limits must be treated as non-negotiable constraints.	Optimization is unsafe if it violates survival margins.
Uncertainty visibility	The system must clearly distinguish facts, estimates, stale data, missing data, and conflicting sensor evidence.	Mars operations will rely on incomplete and delayed information.
Graceful degradation	If data are unreliable or the AI becomes uncertain, the system must recommend conservative procedures or manual fallback.	A weak answer should never be disguised as a strong one.
Traceability	Recommendations, human changes, reasons, rejected alternatives, and outcomes must be recorded.	Delayed mission control review and safety learning require audit trails.

Continued Table 5. *Engineering requirements and guardrails for a future Mars AI decision-support assistant. The table lists candidate design requirements that would help ensure an AI assistant remains explainable, bounded, traceable, and subordinate to human authority. Abbreviation: AI, artificial intelligence.*

Requirement / guardrail	Engineering meaning	Reason for Mars relevance
Planetary protection	The assistant must flag restricted sites, cleanliness rules, sample-handling constraints, and contamination risks.	Scientific integrity and responsible exploration depend on avoiding contamination.
Security and data integrity	Inputs and models must be protected from corruption, misconfiguration, or unauthorized changes.	Bad data or manipulated data could create unsafe recommendations.
Verification and validation	Any operational version must be tested through requirements review, scenario testing, analog exercises, digital twins, and formal safety assurance.	Human-rated use requires evidence, not only plausibility.

SAFETY ASSURANCE AND VALIDATION ROADMAP

For peer-reviewed engineering credibility, the concept must explain how it could later be tested. Autonomy for space missions requires confidence that the system will keep assets safe while helping mission objectives. JPL autonomy-assurance work emphasizes that autonomy creates assurance challenges because future situations cannot always be preplanned, and mission teams need confidence that autonomous behavior will be safe and useful (21). Recent space autonomy research also highlights the importance of mitigating uncertainty, opaqueness, brittleness, and inflexibility while designing human-autonomy teaming interfaces (22).

A safety-assurance pathway should begin with requirements, not code. The team should first identify hazards: unsafe EVA start, loss of suit margin, delayed response to life-support anomaly, robot failure during hazardous inspection, contamination of a sensitive science site, over-reliance on uncertain AI output, and loss of traceability. For each hazard, the team should define what the AI assistant is allowed to do, what it must not do, what information it must show, and when the crew must fall back to manual procedures. Systems-theoretic safety thinking is useful here because accidents can occur not only from component failure, but also from unsafe interactions among humans, software, hardware, procedures, and organizations (23).

The next stage should be low-risk review. Researchers and domain experts can use tabletop scenarios to ask whether the loop helps people notice missing information, question the recommendation, and keep human authority. Later analog missions could test communication delay, crew workload, and habitat-like procedures. A digital twin or operational simulation could then test whether recommendations respect system constraints before any operational claim is made. The purpose of validation is not to show that AI is always right. The purpose is to learn when human-AI teamwork improves awareness and when stronger safeguards are needed.

Figure 3 shows how the concept could mature through increasingly demanding evidence stages. It begins with the conceptual framework itself, then moves to spreadsheet replication or scenario review, sensitivity analysis, tabletop review, analog mission exercises, and finally operational simulation. The sequence is important because human-AI systems should be examined first in settings where errors can teach lessons without endangering people, hardware, or the Martian environment.

Table 6 lays out an evidence ladder for future research. The sequence begins with concept review and tabletop scenarios because low-risk validation should come before analog missions, digital twins, or operational simulations; each stage should provide evidence that the framework improves reasoning, preserves authority, and exposes uncertainty.

RESPONSIBLE AI, PLANETARY PROTECTION AND ETHICS

Responsible AI for Mars should protect astronauts, mission goals, science integrity, and future exploration. Ethical design is not an added topic; it is part of engineering. An assistant that recommends a fast

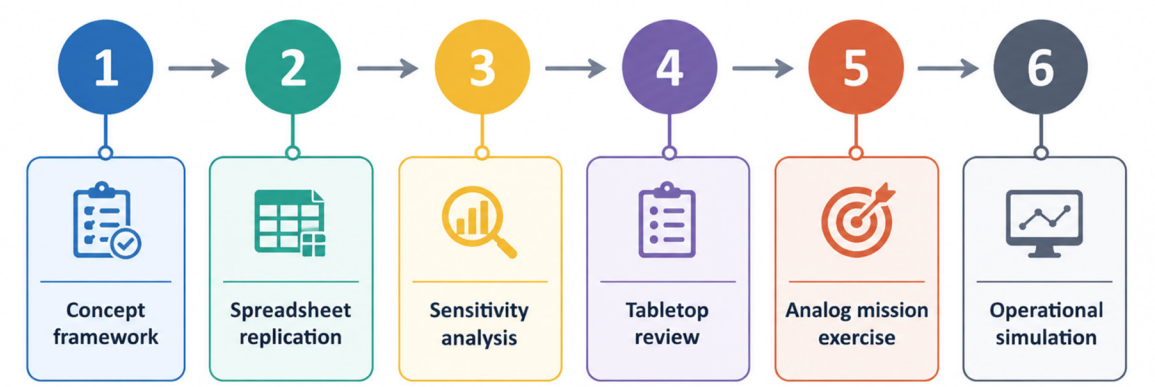

Figure 3. *Staged pathway from concept to possible future validation. The current contribution is the concept framework; later work would need tabletop review, analog exercises, simulation, and formal assurance before any mission use.*

Table 6. *Proposed validation pathway and evidence needed before operational use. The table organizes future research into progressively stronger evidence stages, beginning with concept review and ending with operational qualification; it does not imply that the present paper has completed those stages.*

Stage	Main question	Example evidence	Exit condition
1. Concept review	Is the loop logically clear, scientifically grounded, and ethically responsible?	Expert comments, literature traceability, revised requirements, and scope boundaries.	Reviewers agree the framework is understandable and appropriately limited.
2. Scenario tabletop	Does the loop help reviewers or operators reason through Mars-like situations?	Structured scenarios, decision logs, confusion points, and lessons learned.	Participants can explain recommendations, uncertainties, and human decisions.
3. Human-factors evaluation	Does the interface support situation awareness and calibrated trust?	Measures of comprehension, workload, trust calibration, and ability to detect bad recommendations.	Users neither ignore useful warnings nor over-rely on flawed suggestions.
4. Analog mission exercise	Does the concept help under time delay, workload, and habitat-like constraints?	Analog mission records, communication-delay tests, role observations, and post-run debriefs.	The loop improves coordination without weakening human authority.
5. Digital twin / simulation	Do recommendations obey system constraints and fail safely?	Simulation results, stress cases, anomaly runs, and verification reports.	All safety constraints, fallback procedures, and traceability requirements are satisfied.
Operational qualification	Can a real system meet mission-specific standards and certification needs?	Formal requirements, independent verification, configuration control, cybersecurity review, and mission approval.	Responsible authorities accept the system for limited operational use.

action but ignores contamination risk, crew fatigue, or uncertainty would not be responsible. A useful assistant should show what it knows, what it does not know, why it suggests an action, and what safer alternatives remain available.

Planetary protection is central to Mars exploration. NASA defines planetary protection as the practice of protecting solar system bodies from contamination by Earth life and protecting Earth from possible life forms returned from other solar system bodies (19). The same NASA source emphasizes controlling forward contamination and preventing backward contamination from returned samples. A Mars AI assistant should therefore include planetary-protection constraints as part

of its recommendations, not as an afterthought. It should flag restricted sites, sample-handling rules, clean zones, and records needed for later review.

The framework also supports moral responsibility. If an AI assistant suggests sending a robot into a dangerous area before an astronaut goes outside, the human decision is still meaningful: the crew decides whether the science value justifies the action, whether the robot risk is acceptable, and whether the evidence is strong enough. In this sense, AI does not remove human courage or judgment. It helps make that courage wiser by giving humans a clearer view of risk.

INNOVATION AND SCIENTIFIC CONTRIBUTION

The proposed framework is innovative in three ways. First, it reframes AI for Mars from a tool of task optimization to a tool of shared safety. Efficiency matters, but efficiency without safety, explainability, and planetary protection would be the wrong goal. Second, it treats AI as a translator between Mars and human choice. Mars creates data, warnings, and uncertainty; AI organizes those signals; humans decide what exploration should mean. Third, it links an accessible conceptual figure to engineering ideas: requirements, interfaces, human factors, safety constraints, traceability, and validation.

The strongest version of human-AI teamwork is not full replacement of people by machines. It is disciplined partnership. Robots can go first into uncertain terrain, sensors can watch continuously, AI can connect patterns across data streams, mission control can provide delayed expertise, and astronauts can make final decisions with better awareness. This combination is scientifically and socially important because future exploration must be both brave and responsible.

The framework also has translational value. Non-specialist readers can understand the loop and discuss Mars hazards, AI support, human authority, and ethical safeguards. Researchers and mission designers can extend the same loop into requirements analysis, tabletop scenarios, interface prototypes, simulations, and assurance cases. A good concept should invite deeper research; this one is designed to do exactly that.

LIMITATIONS

The framework has clear limitations. It is conceptual and does not include a trained AI model, mission simulation, astronaut test, analog mission result, or quantitative performance metric. It does not specify hardware architecture, data formats, cybersecurity implementation, medical protocols, or detailed planetary-protection procedures. It also does not prove that AI support will reduce workload or improve safety. Those claims would require future testing.

Another limitation is that human-AI teamwork can introduce new risks. AI may be wrong, overconfident, hard to understand, affected by poor data, or trusted too much by a tired crew. Human operators may ignore correct warnings if they distrust the system or may accept incorrect advice if it appears authoritative. These risks are exactly why the framework emphasizes explainability, human authority, traceability, uncertainty visibility, and staged validation.

CONCLUSION

Mars is too dangerous for humans to explore safely through unaided manual work or delayed Earth support alone. It is also too important to explore through uncontrolled automation. The path between those extremes is human-AI-robot teamwork: machines watch, organize, translate, and assist; humans judge, choose, protect, and learn. Figure 1 captures this idea as a loop linking readiness, in-mission support, human choice, spacecraft architecture, and data-to-value feedback.

The framework developed here is a conceptual contribution, not a tested system. Its value is that it gives future researchers a clear structure for asking the right questions. What is Mars telling us? What does AI think it sees? What is uncertain? What decision must humans make? Which robot or system can act with the least unnecessary risk? What record should be sent to Earth? What did the team learn? These questions are simple, but they are the beginning of engineering discipline.

A smarter path to Mars will require more than powerful rockets and stronger robots. It will require better ways for people and intelligent systems to share awareness, responsibility, and trust. The most inspiring possibility is not that AI explores Mars instead of humans. It is that AI helps humans explore farther, think more clearly, protect life and science, and return safely with knowledge that belongs to all humanity.

ACKNOWLEDGMENTS

The authors thank Dr. Jonathan H. Jiang for mentorship and guidance in developing the research

concept. The authors also thank teachers, reviewers, and future researchers who may continue improving the human-AI teamwork framework for safe and responsible Mars exploration.

DATA AVAILABILITY

The article synthesizes publicly available NASA, JPL, ESA, and scholarly sources cited in the reference list. Future empirical studies should make scenario materials, decision logs, interface mockups, and validation protocols available when appropriate and ethically permissible.

CONFLICT OF INTEREST

The authors declare no conflicts of interest related to this work.

REFERENCES

1. National Aeronautics and Space Administration. Moon to Mars Architecture: strategy and objectives [Internet]. Washington (DC): NASA; 2026 Mar 17. Available from: https://www.nasa.gov/moontomarsarchitecture-strategyandobjectives/ (accessed on 2026-07-03).
2. National Aeronautics and Space Administration. NASA Systems Engineering Handbook. NASA/SP-2016-6105 Rev 2 [Internet]. Washington (DC): NASA; 2016. Available from: https://ntrs.nasa.gov/api/citations/20170001761/downloads/20170001761.pdf (accessed on 2026-07-03).
3. National Aeronautics and Space Administration. Mars: facts [Internet]. NASA Science; 2026 May 26. Available from: https://science.nasa.gov/mars/facts/ (accessed on 2026-07-03).
4. National Aeronautics and Space Administration Human Research Program. 5 hazards of human spaceflight [Internet]. 2026 Feb 27. Available from: https://www.nasa.gov/hrp/hazards/ (accessed on 2026-07-03).
5. National Aeronautics and Space Administration. Environmental Control and Life Support Systems (ECLSS) [Internet]. 2025 Apr 4. Available from: https://www.nasa.gov/reference/environmental-control-and-life-support-systems-eclss/ (accessed on 2026-07-03).
6. European Space Agency. Time delay between Mars and Earth [Internet]. ESA Mars Express Blog; 2012 Aug 5. Available from: https://blogs.esa.int/mex/2012/08/05/time-delay-between-mars-and-earth/ (accessed on 2026-07-03).
7. National Aeronautics and Space Administration Human Research Program. Hazard: distance from Earth [Internet]. 2024 Aug 1. Available from: https://www.nasa.gov/hrp/hazard-distance-from-earth/ (accessed on 2026-07-03).
8. McBrayer KT, Chai PR, Judd EL. Communication delays, disruptions, and blackouts for crewed Mars missions [Internet]. NASA Technical Reports Server; 2022. Available from: https://ntrs.nasa.gov/citations/20220013418 (accessed on 2026-07-03).
9. National Aeronautics and Space Administration. AI inventory [Internet]. Washington (DC): NASA; 2026 May 14. Available from: https://www.nasa.gov/ai-inventory/ (accessed on 2026-07-03).
10. National Aeronautics and Space Administration. NASA's AI use cases: advancing space exploration with responsibility [Internet]. 2025 Jan 7. Available from: https://www.nasa.gov/organizations/ocio/dt/ai/2024-ai-use-cases/ (accessed on 2026-07-03).
11. National Aeronautics and Space Administration Ames Research Center. Planning & Scheduling Group [Internet]. 2023 Dec 14. Available from: https://www.nasa.gov/intelligent-systems-division/autonomous-systems-and-robotics/planning-and-scheduling-group/ (accessed on 2026-07-03).
12. Bresina JL, Jonsson AK, Morris PH, Rajan K. Activity planning for the Mars Exploration Rovers. *Proc Int Conf Automated Planning Scheduling [Internet]*. 2005; 15: 40-49. Available from: https://aaai.org/papers/icaps-05-005-activity-planning-for-the-mars-exploration-rovers/ (accessed on 2026-07-03).
13. Francis R, Estlin T, Doran G, Johnstone S, Gaines D, Verma V, *et al*. AEGIS autonomous targeting for ChemCam on Mars Science Laboratory: deployment and results of initial science team use. *Sci Robot*. 2017; 2 (7): eaan4582. doi:10.1126/scirobotics.aan4582.
14. Jet Propulsion Laboratory. NASA Mars rover can choose laser targets on its own [Internet]. Pasadena (CA): JPL; 2016 Jul 21. Available from: https://www.jpl.nasa.gov/news/nasa-mars-rover-can-choose-laser-targets-on-its-own/ (accessed on 2026-07-03).
15. Jet Propulsion Laboratory. NASA's Perseverance rover completes first AI-planned drive on Mars [Internet]. Pasadena (CA): JPL; 2026 Jan 30. Available from: https://www.jpl.nasa.gov/news/nasas-perseverance-rover-completes-first-ai-planned-drive-on-mars/ (accessed on 2026-07-03).
16. Endsley MR. Toward a theory of situation awareness in dynamic systems. *Hum Factors*. 1995; 37 (1): 32-64. doi:10.1518/001872095779049543.
17. Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. *IEEE Trans Syst Man Cybern A Syst Hum*. 2000; 30 (3): 286-297. doi:10.1109/3468.844354.
18. Lee JD, See KA. Trust in automation: designing for

appropriate reliance. *Hum Factors*. 2004; 46 (1): 50-80. doi:10.1518/hfes.46.1.50_30392.
19. National Aeronautics and Space Administration Office of Planetary Protection. Planetary protection [Internet]. Available from: https://sma.nasa.gov/sma-disciplines/planetary-protection (accessed on 2026-07-03).
20. National Aeronautics and Space Administration. NASA Artificial Intelligence Ethics [Internet]. 2026 Apr 2. Available from: https://www.nasa.gov/nasa-artificial-intelligence-ethics/ (accessed on 2026-07-03).
21. Feather MS, Pinto A. Assurance for autonomy: JPL's past research, lessons learned, and future directions. arXiv. 2023. doi:10.48550/arXiv.2305.11902.
22. Hobbs KL, Phillips S, Simon M, Lyons JB, Culbertson J, Clouse HS, *et al*. The Safe Trusted Autonomy for Responsible Space Program. arXiv. 2025. doi:10.48550/arXiv.2501.05984.
23. Leveson NG. Engineering a safer world: systems thinking applied to safety. Cambridge (MA): MIT Press; 2011. https://doi.org/10.7551/mitpress/8179.001.0001