

A Framework for Advancing Autonomy in Space Robotics: Towards Self-Sustaining Exploration

Filippo Sanfilippo*
University of Agder
Grimstad, Norway
filippo.sanfilippo@uia.no

Connor Esterwood
Wayne State University
Detroit, Michigan
cte@wayne.edu

Tomas Venckunas
Lithuanian Sports University
Kaunas, Lithuania
tomas.venckunas@lsu.lt

Sangseok You
Sungkyunkwan University
Seoul, South Korea
sangyou@skku.edu

Lionel P. Robert Jr.
University of Michigan
Ann Arbor, Michigan
lprobert@umich.edu

Abstract

As humanity advances deeper into space, robotic autonomy offers transformative potential. This paper examines Human-Robot Interaction (HRI) with a focus on how autonomous systems operate in unstructured, extreme environments. We highlight intelligent decision-making for exploration, habitat construction, and resource extraction, drawing parallels between human-led and robot-driven missions. Using the Autonomy Levels for Unmanned Systems (ALFUS) framework, we assess planetary robots across mission complexity (MC), environmental complexity (EC), and external system independence (ESI). The Autonomy and Technology Readiness Assessment (ATRA) method provides a roadmap for progressive capability enhancement. Building on these, we propose Autonomous Robotics for Planetary Exploration (ARPE), a framework defining robotic roles and linking theory to application, and outline future research on long-term robotic sustainability in extreme environments.

Keywords: Autonomous robotics, planetary exploration, space robotics, human-robot interaction.

1. Introduction

As humanity ventures further into space, the challenges of exploration, habitat construction, and resource extraction demand innovative solutions that leverage the strengths of both humans and robots. Central to this endeavour is the concept of robotic autonomy—the ability of a robotic system to perform tasks independently without human input or supervision (Beer et al. (2014)). Robotic autonomy,

underpinned by intelligent decision-making, enables robots to adapt to unexpected and unstructured environments, enhancing their performance in complex scenarios (Hoffman (2019)). In the context of human-robot interaction (HRI), autonomy is crucial, enabling robots to detect, analyse, and respond to human intentions, thereby fostering safer and efficient collaborative systems (Baird and Maruping (2021)).

Autonomous robots in space missions can significantly ease astronauts’ physiological and psychological burdens, especially during long-duration operations. By handling routine tasks, robots allow astronauts to focus on vital countermeasures like exercise, crucial for preventing muscle loss, cardiovascular decline, and stress (Bosutti et al. (2025) and Franchi et al. (2022)). Additionally, robots that support exercise routines—as coaches, partners, or challengers—can boost motivation and adherence through personalised, adaptive feedback (Bosutti et al. (2025)).

However, the successful integration of autonomous robots into space missions hinges on understanding the dynamics of HRI, particularly how varying levels of autonomy influence team coordination, goal alignment, and trust. While prior research has explored these dynamics in terrestrial settings (Chakraborti et al. (2017), Demir et al. (2018), and Johnson et al. (2012)), few studies have considered the unique challenges of extraterrestrial environments, such as communication delays, harsh conditions, and resource scarcity. This gap in the literature underscores the need for a comprehensive framework that integrates robotic autonomy with HRI in space, providing a structured approach to designing collaborative systems that enhance mission efficiency and safety.

This paper re-envision HRI within the context of

Corresponding author. Email: filippo.sanfilippo@uia.no

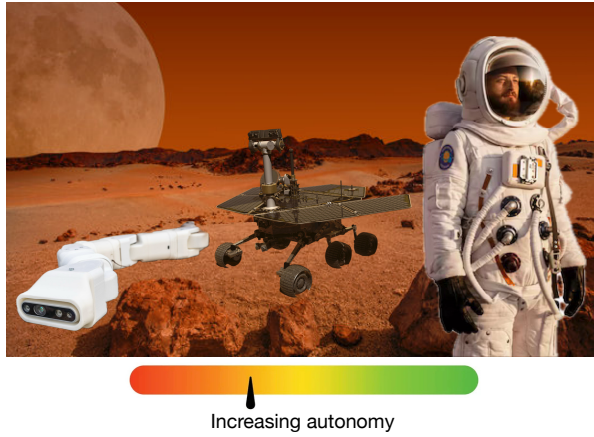


Figure 1. Assessment of planetary robot autonomy and operational interaction in extreme space environments. The underlying idea is to gradually increase autonomy levels for human-robot interaction (HRI) in off-world missions.

planetary exploration, focusing on how autonomous robotic systems and their human counterparts interact with dynamic, unpredictable terrains while facing the numerous challenges presented by limited communication. We analyse the role of autonomy in robotic exploration and its impact on mission adaptability, drawing comparisons between traditional human-led space missions and autonomous planetary operations, as shown in Fig. 1. By examining the effects of increasing autonomy, we aim to advance robotic exploration through a systematic study of self-sufficiency. Similar to how human astronauts rely on mission control for strategic guidance, autonomous planetary robots initially require remote supervision. As autonomy increases, they transition to self-directed operations, adapting navigation routes, responding to unforeseen hazards, and executing scientific tasks independently. Ground control assumes the role of high-level mission planners, ensuring that robots learn from past missions and refine their operational efficiency. Understanding these parallels can provide insights into the development of next-generation autonomous planetary systems and their integration into long-term space missions.

To ensure a structured methodology, the Autonomy Levels for Unmanned Systems (ALFUS) framework (Huang et al. (2005)) is applied to evaluate the autonomy of planetary robots, considering mission complexity (MC), environmental complexity (EC), and external system dependence (ESI). Additionally, the Autonomy and Technology Readiness Assessment (ATRA) methodology is utilised to incrementally

develop robotic capabilities, ensuring a systematic progression toward higher autonomy while addressing both technological and operational challenges.

This work introduces a conceptual framework for planetary exploration, termed *Autonomous Robotics for Planetary Exploration* (ARPE). The framework defines and contextualises the roles of key robotic agents, leveraging established autonomy frameworks to provide a structured approach to increasing autonomy in planetary robotics. This research aims to provide a step-wise approach for enhancing planetary robotic guidance, navigation, and control (GNC) capabilities, targeting various autonomy levels. By doing so, the paper not only contributes to the evolution of autonomous space exploration but also advances HRI research by exploring the interaction dynamics between autonomous robotic systems and extreme planetary environments.

The paper is structured as follows. A review of related research is provided in Section 2. Section 3 discusses the unique challenges of planetary robotic operations and the role of autonomous systems. In Section 4, we apply the ALFUS framework to analyse autonomy within planetary missions. Section 5 explores the ATRA methodology and its implications for developing self-sufficient planetary robots. Based on these methodologies, the *Autonomous Robotics for Planetary Exploration* (ARPE) framework is introduced. In Section 6, human factors are described. Section 7 presents a case study evaluating the framework’s applicability in planetary robotic operations. A discussion is outlined in Section 8. Conclusions and future directions are discussed in Section 9.

2. Related Works

2.1. Robot Autonomy

Robot autonomy—the capacity to perform tasks without human intervention—is driven by integrated perception, decision-making, and control (Beer et al. (2014)). It enables adaptive behavior in unstructured environments, improving robot performance in complex settings (Goodrich and Schultz (2007) and Hoffman (2019)). In HRI, autonomy enhances safety and collaboration by enabling robots to interpret and respond to human intent (Baird and Maruping (2021) and Nikolaidis et al. (2017)).

In construction, autonomous systems improve productivity and safety by handling repetitive and hazardous tasks (Arashpour et al. (2016), Delgado et al. (2019), and You et al. (2018)). Similarly, in space

exploration, autonomy addresses communication delays and environmental challenges. NASA's Perseverance rover exemplifies this by autonomously navigating, sampling, and making strategic decisions (Verma et al. (2023) and Washington et al. (1999)).

Recent work integrates machine learning and virtual environments to expand autonomy (Verma et al. (2023)), yet challenges remain for robust performance in unstructured, remote settings. This is critical as autonomy underpins essential space tasks like habitat construction and resource extraction. A structured framework accounting for autonomy levels, system independence, and environmental complexity is vital for designing resilient, self-sustaining robots in planetary exploration.

2.2. Humans In Space

Extended space missions pose physiological and psychological risks, particularly in microgravity. These include vestibular dysfunction impairing balance and motor coordination, appetite loss, muscle atrophy, and cardiovascular deconditioning (Bosutti et al. (2025) and Franchi et al. (2022)). Bone mineral density also decreases, increasing fracture risk upon gravity restoration (Minetti et al. (2024)). Additional maladaptations include fluid shifts causing orthostatic intolerance and reduced blood volume, alongside metabolic issues like insulin resistance and mitochondrial dysfunction (Bosutti et al. (2025) and Murgia et al. (2022)). Psychologically, confinement and reduced interaction cause mood disturbances, especially in active individuals (Sigita et al. (2024)). While social robots may alleviate isolation, the stress of confinement persists.

Robotic autonomy offers a promising solution. As autonomy increases, robots can manage routine tasks, freeing astronauts for vital countermeasures like exercise (Bosutti et al. (2025) and Franchi et al. (2022)). Robots could also augment physical training as adaptive partners, delivering feedback and implementing electrical stimulation protocols. These roles reduce cognitive load and support physical and mental well-being. By embedding robots into daily life and adapting to individual needs, autonomous systems can enhance mission sustainability and astronaut resilience during long-duration exploration.

3. Key Challenges in Planetary Autonomy

Advancing planetary robot autonomy involves tackling challenges beyond typical robotics. These robots must function in remote, unstructured, and extreme environments with minimal human

oversight and significant communication delays, making autonomous operation essential.

Planetary surfaces are unpredictable—featuring rugged terrain, temperature extremes, radiation, and dust storms. Robots must adapt to these hazards dynamically, unlike terrestrial systems which operate in controlled conditions.

Communication delays between Earth and space prohibit real-time control. Robots must independently assess surroundings, navigate, and manage emergencies using local sensor data, ensuring mission continuity without ground input.

Limited power and computing resources further constrain autonomy. Space hardware lags behind terrestrial systems in performance, and power—often solar—is scarce. Efficient, low-power algorithms are required for reliable autonomous operations.

Future missions will likely rely on multi-robot teams for tasks like habitat construction and resource extraction. These robots must collaborate without human input, dynamically adapting to evolving objectives and conditions. Sophisticated task-allocation and cooperation algorithms are needed to manage such teams.

Robots must also detect and recover from failures autonomously. With no human repair possible, onboard fault detection and recovery are critical for sustained operations.

Though direct HRI is constrained in space, human operators still rely on interfaces to understand robot behaviour and make informed decisions. Trustworthy, transparent decision-making and intuitive feedback mechanisms are vital. Robots must explain autonomous decisions clearly to build operator trust and enable seamless cooperation, even across long distances.

A key aspect of advancing autonomy in these systems is trust and adaptability. Robotic systems must gradually transition from simple, human-supervised tasks to complete self-sufficiency, with each step thoroughly tested in Earth-based analogues to ensure reliability. This progression is supported by frameworks like the Autonomy Levels for Unmanned Systems (ALFUS) and the Autonomy and Technology Readiness Assessment (ATRA), which offer structured pathways to enhance autonomy in planetary robots. These frameworks evaluate the complexity of the mission, the environment, and the degree of independence required, allowing for systematic improvements and risk management.

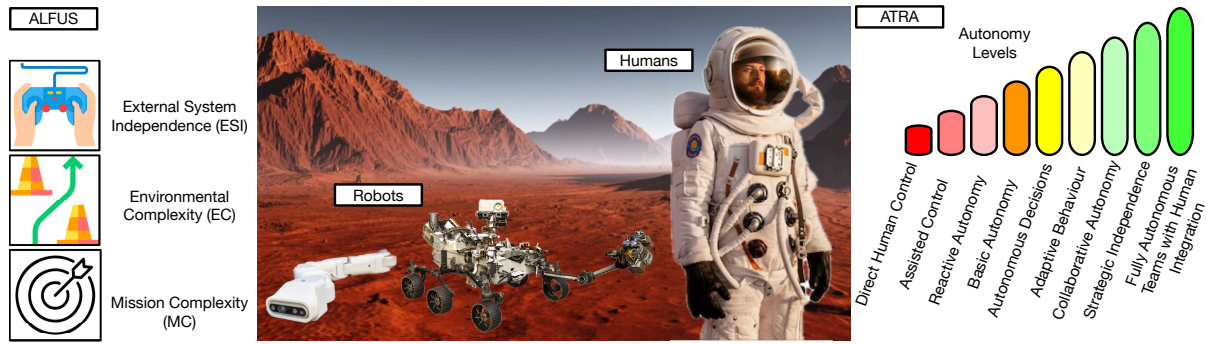


Figure 2. The proposed framework, named the Autonomous Robotics for Planetary Exploration (ARPE), is based on the application of the ALFUS and ATRA methods to planetary robots.

4. Autonomy Levels for Unmanned Systems (ALFUS) in Planetary Robotics

When designing autonomous planetary robots, it is crucial to assess their operational autonomy across several dimensions. These dimensions include environmental complexity (EC), which refers to the ability of the robots to navigate and function within dynamic, highly uncertain planetary environments, such as rough terrains, extreme weather conditions, or unknown geological formations; mission complexity (MC), which involves the range and intricacy of tasks the robots must perform, such as conducting scientific experiments, gathering samples, navigating unpredictable landscapes, and collaborating with other robots or systems; and external system independence (ESI), which measures the degree to which these robots can operate autonomously without relying on Earth-based control, communication, or external human intervention.

Based on these metrics, the *Autonomy Levels For Unmanned Systems* (ALFUS) framework (Huang et al. (2005)) can be adopted and applied to provide a more in-depth overview for the design of autonomous planetary robots, as shown in Figure 2. Regarding external system independence (ESI), a spectrum of autonomy levels (ALs) can be defined as shown in Table 1. Concerning environmental complexity (EC), different ALs can be identified according to gradually increased complexity, as shown in Table 2. With reference to mission complexity (MC), various ALs can be outlined according to gradually increased complexity, as shown in Table 3.

The application of the ALFUS framework in planetary robotics enables a clear roadmap for the gradual increase in autonomy required for space missions. By evaluating mission complexity,

environmental complexity, and external system independence, the ALFUS levels help define the stages of robotic capability development, from basic teleoperation to fully autonomous systems capable of self-sustaining operations. The framework serves as a critical tool in guiding the design and testing of planetary robots, ensuring that each robot reaches its full potential in terms of both autonomy and mission success.

In conclusion, the ALFUS framework provides a structured approach to enhancing planetary robotic autonomy. As robots advance through the autonomy levels, they will be able to perform increasingly complex tasks independently, ensuring the success of long-duration space missions where human involvement is limited or delayed. The progression through these levels will enable robotic systems to meet the growing demands of space exploration, ultimately contributing to humanity's ability to sustain life on other planets.

5. Autonomy and Technology Readiness Assessment (ATRA) and HRI in Planetary Robotics

To systematically evaluate the Autonomy Levels (ALs) of planetary robots deployed for space exploration, the *Autonomy and Technology Readiness Assessment* (ATRA) framework (Kendoul (2013)) may be adopted, as illustrated in Fig.2. The ATRA framework integrates both autonomy levels and Technology Readiness Level (TRL) metrics to provide a holistic evaluation. Inspired by its application in similarly complex robotic systems, such as snake robots (Sanfilippo, Wiley, and Rousi (2025) and Sanfilippo et al. (2017)), this framework can be employed to deliver a comprehensive understanding of how planetary robotic systems can progressively achieve mission objectives. In the following, a nine-level scale framework is proposed based on the

Table 1. Autonomy Levels (AL) of Planetary Robots Based on External System Independence (ESI)

AL	ESI	Description
1	High Dependence	Robots fully reliant on Earth-based control; no autonomy, functioning as remote-controlled devices.
2	High Dependence	Limited autonomy; robots assist with basic functions like navigation or data collection, still requiring significant human control.
3	Moderate Dependence	Robots can execute predefined tasks autonomously when prompted, but rely on Earth-based systems for complex decisions or emergency response.
4	Moderate Dependence	Robots handle routine tasks autonomously but require external help for unexpected or strategic scenarios.
5	Low Dependence	Robots make autonomous decisions in controlled environments, relying minimally on Earth-based systems for high-level oversight.
6	Low Dependence	Adapt to dynamic planetary environments with occasional Earth-based input, managing complex mission strategies autonomously.
7	Very Low Dependence	Mostly autonomous, collaborating with humans during critical moments, needing minimal external support.
8	Minimal Dependence	Independently manage chaotic planetary environments, consulting Earth-based systems primarily for post-mission analysis.
9	Complete Independence	Fully autonomous, operating independently of Earth control, learning and adapting in real-time without external input.

gradual increase (autonomy as a gradual property) of guidance, navigation, and control (GNC) functions and capabilities. The proposed framework is named

Table 2. Autonomy Levels (AL) of Planetary Robots Based on Environmental Complexity (EC)

AL	EC	Description
1	Static	Simple, controlled environments like indoor labs or flat, stable surfaces; robots perform basic tasks with no dynamic elements.
2	Slightly Dynamic	Environments with minor changes; robots manage basic navigation with small, predictable obstacles such as rocks or debris.
3	Controlled Dynamic	Environments with predictable dynamics such as slow-moving terrain or minor obstacles; robots begin to react to these changes.
4	Moderate Dynamic	Environments that are more varied with multiple dynamic elements, such as uneven terrain or slight environmental changes like wind or temperature.
5	Highly Dynamic	Complex, rapidly changing environments with unpredictable factors like dust storms or terrain shifts; robots need to adapt continuously to new conditions.
6	Semi Unpredictable	Highly dynamic environments with semi-unpredictable factors like varying terrain and potential obstacles; robots must anticipate and adapt.
7	Unpredictable	Environments with frequent, sudden changes and unknowns; robots must make real-time decisions in complex, unforeseen scenarios.
8	Chaotic	Environments with chaotic dynamics, such as dust storms or rapidly shifting landscapes; robots operate with high adaptability and real-time strategy adjustments.
9	Extremely Chaotic	The highest level of environmental complexity; robots autonomously manage extreme unpredictability and continuous environmental changes, such as planetary quakes or massive dust clouds.

Autonomous Robotics for Planetary Exploration (ARPE).

Level 1: Fully Dependent on Human Control. At

Table 3. Autonomy Levels (AL) of Planetary Robots Based on Mission Complexity (MC)

AL	MC	Description
1	Basic Movements	Robots perform simple actions such as basic movement, obstacle avoidance, or simple sample collection without specific mission goals.
2	Simple Task Execution	Robots complete straightforward tasks like moving to a target location or taking a single set of measurements.
3	Basic Interaction	Robots interact with the environment in simple, predictable ways, such as collecting samples or performing environmental scans.
4	Coordinated Task Execution	Robots perform coordinated actions, such as working in tandem to gather samples or navigate together.
5	Tactical Play	Robots begin to implement simple mission strategies, such as prioritising objectives or adapting to basic environmental shifts.
6	Dynamic Role Adaptation	Robots dynamically adjust their roles during the mission based on environmental conditions or shifting objectives, such as switching between exploration and resource extraction.
7	Complex Strategy	Robots execute complex, multi-step mission strategies, such as coordinating with other robots or responding to unforeseen environmental changes.
8	Autonomous Mission Management	Robots autonomously manage all aspects of the mission, including real-time adaptation of strategies and task re-prioritization.
9	Strategic and Long-Term Planning	Robots not only manage immediate tasks but also engage in long-term mission planning, learning from past experiences and adapting strategies over time.

the lowest level of autonomy, robots are fully dependent on human operators. They function as remote-controlled systems, with no decision-making capability of their own. Human operators guide every action, from navigation to task execution, and the robot responds in real-time to external commands. This level of autonomy is typical during early stages of mission testing or when robots are deployed in environments where human supervision is required for safety reasons.

Level 2: Limited Autonomy with Human Assistance. At this level, robots begin to execute predefined tasks autonomously but still rely heavily on human oversight. They may perform basic functions such as moving along a specific route or carrying out simple tasks like taking images or collecting samples. However, more complex decision-making, such as navigating obstacles or responding to unexpected environmental changes, requires human intervention. This level of autonomy is useful for missions in controlled or partially unpredictable environments where safety and reliability are paramount.

Level 3: Moderate Autonomy with Supervisory Control. In this stage, robots can handle more complex tasks autonomously, such as terrain analysis, sample collection, or simple navigation adjustments. However, they still rely on external guidance for complex decision-making. The robot may react to some environmental changes autonomously, such as avoiding obstacles or adjusting speed based on terrain conditions, but more advanced tasks—such as re-planning a route in response to a sudden hazard—require human supervision. This level is suitable for missions where robots need to perform basic autonomous actions but still need frequent input from mission control.

Level 4: High Autonomy with Occasional Human Intervention. At this level, robots can perform a wide range of tasks independently. They are capable of dynamic decision-making, such as adjusting their path to avoid unforeseen obstacles or adapting to changes in the environment, like shifting terrain or temperature variations. Human intervention is only required in rare, complex situations, such as emergency decision-making or significant re-planning. This level of autonomy is essential for longer-duration missions, where robots must work autonomously in challenging environments, but still retain the option for remote oversight when needed.

Level 5: Autonomous with Minimal Human Involvement. Level 5 robots exhibit a high degree of autonomy, capable of carrying out most mission tasks without human input. They can handle complex environmental factors, such as navigating through rough terrain, adjusting to temperature shifts, and reacting to

unexpected events autonomously. Humans are involved only in high-level mission planning and intervention in cases of major system failure or critical mission changes. This level is critical for missions in remote locations, where continuous human communication is impossible or impractical.

Level 6: Highly Autonomous with Limited Human Interaction. Robots at this level of autonomy are capable of managing all aspects of their mission with minimal human involvement. They can adapt to highly dynamic environments, make complex decisions in real time, and coordinate with other robotic agents on multi-robot tasks. Human intervention is almost entirely limited to strategic decision-making or adjustments in mission parameters. These robots are suited for missions in extremely remote or hostile environments, such as Mars or deep space exploration, where human interaction is limited to pre-deployment planning and post-mission analysis.

Level 7: Fully Autonomous with Collaborative Capabilities. At this level, robots are fully autonomous and can collaborate with other robots and humans. They can manage complex mission objectives, interact with environmental elements, and adjust strategies based on real-time data. These robots are capable of performing advanced tasks such as autonomous excavation, habitat construction, and scientific research without human assistance. Human interaction, if needed, is only for high-level coordination and oversight. The ability to collaborate with other autonomous systems makes this level ideal for missions involving multi-robot teams working together to achieve common goals.

Level 8: Near-Complete Autonomy with Self-Optimization. Level 8 robots are capable of near-total autonomy. They operate independently across a wide range of environmental conditions and mission types, utilising advanced learning algorithms to optimise their performance over time. These robots can self-monitor, identify areas for improvement, and adjust their behaviours to enhance mission success. Human involvement is minimal, typically limited to post-mission data analysis or high-level strategic planning. These robots are suitable for missions that require extended durations and high levels of adaptability, such as long-term planetary exploration or asteroid mining.

Level 9: Fully Autonomous with Self-Sustaining Capabilities. The highest level of autonomy is achieved when robots can operate entirely independently, without any need for human oversight or intervention. They can manage all aspects of the mission, adapt to dynamic environments, interact with other robotic agents, and make decisions based on a combination of

pre-programmed goals and real-time data. These robots are capable of performing complex, long-duration missions that may involve maintaining themselves, repairing minor faults, and optimising their systems for extended operational life. Level 9 autonomy represents the ultimate goal for planetary robotics, enabling fully self-sustaining exploration systems that can operate on distant planets or moons for extended periods, contributing to humanity's long-term space exploration ambitions.

The ARPE framework currently assumes autonomy levels rise equally across ESI, EC, and MC. In reality, these dimensions may progress independently—e.g., high ESI can occur with moderate EC. Instead of a linear path, autonomy should be viewed as a multidimensional space with varying combinations shaped by context. Adopting this flexible approach would enable ARPE to better capture diverse space mission scenarios while maintaining structured autonomy assessment.

6. Human Factors

The ARPE framework shows strong potential, but varying autonomy levels introduce challenges for HRI. Even highly autonomous robots require interaction, making it crucial to examine autonomy's impact on coordination, goal alignment, and trust. Existing research in this area is limited. This section outlines key challenges, provide recommendations where possible, and identify areas for future research.

Team coordination involves “strategies and patterns of behaviour aimed to integrate actions, knowledge and goals of interdependent members to achieve common goals” (Rico et al. (2011, Pg.59)) and is a key factor in team performance (Demir et al. (2018)). While human-human teams rely on dynamic coordination, human-robot teams often use more rigid methods (Demir et al. (2018)). As autonomy increases, coordination may become more flexible and adapt to complex goals and environments. However, how variable autonomy affects coordination in human-robot teams remains unclear (Chakraborti et al. (2017), Demir et al. (2018), Johnson et al. (2012), and Sellner et al. (2006)). A comprehensive synthesis is lacking, and existing work rarely addresses challenges specific to space operations. Future studies should explore these gaps, using the ARPE framework to better understand autonomous coordination in extraterrestrial HRI.

Goal alignment is the extent to which an agent's goals correspond with those of its human counterpart (Lee (2024), Li and Lee (2022), Salas et al. (1995), and Salehzadeh et al. (2022)), enabling efficient collaboration (Lee (2024) and Muggleton and

Chater (2021)) and reducing conflict and mission risk when well-aligned (Li and Lee (2022)). Within the ARPE framework, maintaining alignment becomes increasingly challenging at higher autonomy levels as decision-making grows more abstract, opaque, and risky (Li and Lee (2022), Lyons et al. (2021), Tallam (2025), and Vaassen (2022)). Highly autonomous systems pose a “significantly higher risk of misalignment and instrumental convergence” (Tallam (2025, Pg.19)). Approaches for aligning human and autonomous system goals exist (Lee (2024), Li and Lee (2022), and Tallam (2025)) and, to a lesser extent, for HRI (Zhang et al. (2025)). Notably, (Tallam (2025)) parallels ARPE, with Level 0 (“Reactive”) matching lower autonomy and Level 3 (“Self-Directed”) higher autonomy (Tallam (2025, p.21)). These principles could inform HRI in space operations and warrant further study.

Trust is defined as a willingness to be vulnerable based on the belief that another will act beneficially (Mayer et al. (1995, Pg.712)) and is critical for effective collaboration (Lyons et al. (2019), Robert (2020), Wiese et al. (2017), and You and Robert (2018a)), including in HRI (Esterwood and Robert (2020), Hancock et al. (2011, 2021), and You and Robert (2018b)). Unlike humans, robots operate across variable autonomy levels, complicating trust dynamics. Many trust models in HRI assume static autonomy (Beer et al. (2014), Firmino de Souza et al. (2025), and Law and Scheutz (2021)), leaving it unclear how shifts in autonomy, as seen in ARPE, affect trust development and stability. While levels of automation—defined as degrees of system control (Parasuraman et al. (2000))—have been more explored, results are mixed, showing no clear link between automation and trust (Rani et al. (2000) and Schaefer et al. (2016)). As robots gain adaptive autonomy, deeper research is needed to understand how trust evolves across dynamic autonomy levels. Without this, systems risk being under- or over-trusted, leading to inefficiency or unsafe behavior. Future work should study how users perceive autonomy changes, calibrate trust accordingly, and how design can support reliable trust across ARPE levels.

7. Mars Colonization Use Case

In the context of future Mars missions, effective HRI will be vital to establishing sustainable habitats and supporting long-term exploration objectives. Robotic systems equipped with advanced functionalities—such as robotic arms with integrated 3D printing capabilities—will autonomously fabricate habitats and structures utilising local regolith, thus reducing

dependency on Earth-supplied materials (Cesaretti et al. (2014) and Netti (2025)). These construction robots will operate under Environmental Complexity (EC) Autonomy Levels 3-4, wherein they respond to slow terrain shifts and surface irregularities, executing coordinated construction tasks with minimal human intervention. Meanwhile, higher-level tactical planning, including complex task coordination and environmental assessment, will be managed at EC Levels 5-6 (Mission Complexity), enabling adaptive responses to dynamic surface conditions and supporting construction (Tamminen et al. (2016)).

Astronauts will oversee these processes and intervene when tasks require precision, supported by autonomous systems that adapt to contextual demands. To ensure mission safety and foster trust, spatial separation between human workspaces and robotic operations will be maintained through geo-fenced or physical boundaries, aligning with the principles of the Robot Acceptance Safety Model (RASM). Such spatial delimitation encourages team identification and trust, which are crucial for mitigating perceived risks and fostering continued cooperation with robotic systems (Cabrol (2018) and You et al. (2018)). Maintaining delineated work zones not only enhances psychological safety but also supports a broad spectrum of collaborative tasks, fostering effective HRI (You et al. (2018)).

Beyond construction activities, robotic systems will undertake resource extraction and environmental management tasks, such as subsurface ice harvesting for water and oxygen production. Additionally, robots will perform autonomous self-diagnostics, routine maintenance, and soil analysis for biosignatures—activities essential for supporting long-duration missions under extreme conditions (You and Robert Jr (2024)). These systems are expected to operate at higher levels of autonomy—EC Levels 6-7—which encompass semi-unpredictable to unpredictable environmental scenarios—and require Mission Complexity (MC) Levels 6-7, involving dynamic role adaptation and complex strategy implementation (Tamminen et al. (2016)). Such autonomy enables real-time role switching and decision-making based on environmental cues, ensuring sustainability and operational resilience.

In addition, for Mars colonisation, autonomous robots will play a key role in:

- **Habitat Construction:** 3D printing technology and robotic arms will build shelters using in-situ materials (EC Autonomy Levels 3-4).
- **Resource Mining:** Robots will extract ice

for water and oxygen generation, supporting long-term human missions (EC and MC Autonomy Levels 6-7).

- **Maintenance and Repair:** Self-repairing robotic systems will enhance mission sustainability by detecting and fixing malfunctions (EC Autonomy Levels 5–6 and MC Autonomy Levels 7-8).
- **Scientific Exploration:** Advanced AI systems will analyse soil samples and search for signs of past or present microbial life (EC and MC Autonomy Levels 7-8).

This use case highlights the critical role of robotic autonomy in enabling Mars colonization, demonstrating how intelligent decision-making and adaptive interaction can enhance mission efficiency and safety. By drawing parallels between human-led and robot-driven tasks, this paper provides a framework that links robot autonomy to HRI, advancing the discourse on collaborative systems in space exploration.

The proposed framework not only addresses the immediate challenges of Mars colonization but also sets the stage for future research and applications, ensuring the success and sustainability of long-duration planetary missions. By integrating autonomy into the broader context of HRI, this work underscores the transformative potential of robotics in advancing humanity's exploration of space.

8. Discussion

The ARPE framework addresses a gap in current literature by offering a structured model for human-robot collaboration in space missions, including exploration, habitat construction, and resource extraction.

In team coordination, this work builds on insights from (Rico et al. (2011)) by showing how integrated actions, knowledge, and goals improve shared outcomes. Unlike human teams that coordinate flexibly, robotic teams often depend on rigid methods (Demir et al. (2018)). ARPE enables more adaptive coordination by incorporating autonomy levels to accommodate communication delays and environmental constraints in space.

Regarding goal alignment, ARPE highlights the necessity of matching robotic goals with human intentions (Lee (2024), Li and Lee (2022), Salas et al. (1995), and Salehzadeh et al. (2022)), as misalignment can cause inefficiencies and risk (Tallam (2025)). Higher autonomy levels increase decision opacity (Li and Lee (2022), Lyons et al. (2021), and Vaassen (2022)), demanding mechanisms to maintain alignment.

ARPE aligns with established taxonomies, such as Level 0 (reactive) to Level 3 (self-directed) systems (Tallam (2025)), adapting goal-alignment principles to space HRI.

In terms of trust, this work responds to the lack of models addressing trust under changing autonomy levels (Beer et al. (2014), Firmino de Souza et al. (2025), and Law and Scheutz (2021)). ARPE expands on prior models (Rani et al. (2000) and Schaefer et al. (2016)) by examining trust dynamics in adaptive systems, which are critical for safe and efficient astronaut-robot collaboration.

The framework also encourages testing in harsh terrestrial analogue environments such as Arctic expeditions, which offer realistic conditions for simulating Martian HRI (Johnsen et al. (2021)).

Finally, while tailored for space, ARPE applies to other domains—e.g., disaster response (Sanfilippo, Moosavi, and Zafar (2025)), deep-sea exploration, industrial settings, and eldercare robotics (Sanfilippo, Fossum, and Hansen (2025))—by supporting robust HRI in unpredictable environments.

9. Conclusion

This paper examined the dynamics of human-robot interaction (HRI) in space exploration, focusing on how autonomous systems can support mission tasks while fostering trust and coordination among crew. Using the Autonomy Levels for Unmanned Systems (ALFUS) framework, we assessed autonomy through mission complexity (MC), environmental complexity (EC), and external system independence (ESI). The Autonomy and Technology Readiness Assessment (ATRA) method provided a roadmap for improving robot capabilities and operational readiness. Based on these, we introduced the *Autonomous Robotics for Planetary Exploration* (ARPE) model, with a case scenario illustrating its application. Future work should explore real-time trust metrics, adaptive interfaces aligned with human states, and explainable autonomy in mission-critical settings. Extended Reality (XR) technologies and multimodal cues may further enhance interaction fluency and human-robot mutual understanding.

Acknowledgement

This research is supported by the Artificial Intelligence, Biomechatronics, and Collaborative Robotics research group at the Top Research Center Mechatronics (TRCM), University of Agder (UiA), Norway, the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea

(NRF-2023S1A5A8081097).

References

- Arashpour, M., Wakefield, R., Lee, E., Chan, R., & Hosseini, M. R. (2016). Analysis of interacting uncertainties in on-site and off-site activities: Implications for hybrid construction. *International Journal of Project Management*, 34(7), 1393–1402.
- Baird, A., & Maruping, L. M. (2021). The next generation of research on is use: A theoretical framework of delegation to and from agentic is artifacts. *MIS quarterly*, 45(1).
- Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of human-robot interaction*, 3(2), 74.
- Bosutti, A., Ganse, B., Maffiuletti, N. A., Wüst, R. C., Strijkers, G. J., Sanderson, A., & Degens, H. (2025). Microgravity-induced changes in skeletal muscle and possible countermeasures: What we can learn from bed rest and human space studies. *Experimental Physiology*.
- Cabrol, N. A. (2018). The coevolution of life and environment on mars: An ecosystem perspective on the robotic exploration of biosignatures.
- Cesaretti, G., Dini, E., De Kestelier, X., Colla, V., & Pambaguian, L. (2014). Building components for an outpost on the lunar soil by means of a novel 3d printing technology. *Acta Astronautica*, 93, 430–450.
- Chakraborti, T., Kambhampati, S., Scheutz, M., & Zhang, Y. (2017). Ai challenges in human-robot cognitive teaming. *arXiv preprint arXiv:1707.04775*.
- Delgado, J. M. D., Oyedele, L., Ajayi, A., Akanbi, L., Akinade, O., Bilal, M., & Owolabi, H. (2019). Robotics and automated systems in construction: Understanding industry-specific challenges for adoption. *Journal of building engineering*, 26, 100868.
- Demir, M., Likens, A. D., Cooke, N. J., Amazeen, P. G., & McNeese, N. J. (2018). Team coordination and effectiveness in human-autonomy teaming. *IEEE Transactions on Human-Machine Systems*, 49(2), 150–159.
- Esterwood, C., & Robert, L. P. (2020). Human robot team design. *Proc. of the 8th International Conference on Human-Agent Interaction*, 251–253.
- Firmino de Souza, D., Sousa, S., Kristjhan-Ling, K., Dunajeva, O., Roosileht, M., Pentel, A., Möttus, M., Can Özdemir, M., & Gratsjova, Ž. (2025). Trust and trustworthiness from human-centered perspective in human-robot interaction (hri)—a systematic literature review. *Electronics*, 14(8), 1557.
- Franchi, M. V., Sarto, F., Simunič, B., Pišot, R., & Narici, M. V. (2022). Early changes of hamstrings morphology and contractile properties during 10 d of complete inactivity. *Medicine and Science in Sports and Exercise*, 54(8), 1346–1354.
- Goodrich, M. A., & Schultz, A. C. (2007). Human-robot interaction: A survey, foundations and trends in human-computer interaction, vol. 1.
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, 53(5), 517–527.
- Hancock, P. A., Kessler, T. T., Kaplan, A. D., Brill, J. C., & Szalma, J. L. (2021). Evolving trust in robots: Specification through sequential and comparative meta-analyses. *Human factors*, 63(7), 1196–1229.
- Hoffman, G. (2019). Evaluating fluency in human-robot collaboration. *IEEE Transactions on Human-Machine Systems*, 49(3), 209–218.
- Huang, H.-M., Pavek, K., Novak, B., Albus, J., & Messin, E. (2005). A framework for autonomy levels for unmanned systems (alfus). *Proc. of the AUVSI's unmanned systems North America*, 849–863.
- Johnsen, B. H., Brattebø, G., Phillips, T. M., Gjeldnes, R., Bartone, P. T., Monsen, H.-O. N., & Thayer, J. F. (2021). Crossing the antarctica: Exploring the effects of appetite-regulating hormones and indicators of nutrition status during a 93-day solo-expedition. *Nutrients*, 13(6), 1777.
- Johnson, M., Bradshaw, J. M., Feltoovich, P., Jonker, C., Van Riemsdijk, B., & Sierhuis, M. (2012). Autonomy and interdependence in human-agent-robot teams. *IEEE Intelligent Systems*, 27(2), 43–51.
- Kendoul, F. (2013). Towards a unified framework for uas autonomy and technology readiness assessment (atra). *Autonomous Control Systems and Vehicles: Intelligent Unmanned Systems*, 55–71.
- Law, T., & Scheutz, M. (2021). Trust: Recent concepts and evaluations in human-robot interaction. *Trust in human-robot interaction*, 27–57.
- Lee, J. (2024). *Influence of goal alignment on delegation decisions to human and automated collaborators* [Doctoral dissertation, Arizona State University] [Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2024-04-18].
- Li, M., & Lee, J. D. (2022). Modeling goal alignment in human-ai teaming: A dynamic game theory approach. *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, 66(1), 1538–1542.
- Lyons, J. B., Sycara, K., Lewis, M., & Capiola, A. (2021). Human-autonomy teaming: Definitions, debates, and directions. *Frontiers in psychology*, 12, 589585.
- Lyons, J. B., Wynne, K. T., Mahoney, S., & Roebke, M. A. (2019). Trust and human-machine teaming: A qualitative study. In *Artificial intelligence for the internet of everything* (pp. 101–116). Elsevier.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, 20(3), 709–734.
- Minetti, A. E., Luciano, F., Natalucci, V., & Pavei, G. (2024). Horizontal running inside circular walls of moon settlements: A comprehensive countermeasure for low-gravity deconditioning? *Royal Society Open Science*, 11(5), 231906.
- Muggleton, S., & Chater, N. (2021). *Human-like machine intelligence*. Oxford University Press. <https://books.google.com/books?id=wX0xEAAQBAJ>
- Murgia, M., Ciciliot, S., Nagaraj, N., Reggiani, C., Schiaffino, S., Franchi, M. V., Pišot, R., Simunič, B., Toniolo, L., Blaauw, B., et al. (2022). Signatures of muscle disuse in spaceflight and bed rest revealed by single muscle fiber proteomics. *PNAS nexus*, 1(3), pgac086.
- Netti, V. (2025). Planetary construction with isru on moon and mars: Scalability and automation of the construction processes.
- Nikolaïdis, S., Hsu, D., & Srinivasa, S. (2017). Human-robot mutual adaptation in collaborative tasks: Models and experiments. *The International Journal of Robotics Research*, 36(5-7), 618–634.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 30(3), 286–297.
- Rani, M. A., Sinclair, M. A., & Case, K. (2000). Human mismatches and preferences for automation. *International Journal of Production Research*, 38(17), 4033–4039.
- Rico, R., Sánchez-Manzanares, M., Gil, F., Maria Alcover, C., & Tabernero, C. (2011). Coordination processes in work teams. *Papeles del Psicólogo*, 32(1), 59–68.
- Robert, L. P. (2020). Behavior-output control theory, trust and social loafing in virtual teams. *Multimodal Technologies and Interaction*, 4(3), 39.
- Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human factors*, 37(1), 123–136.
- Salehzadeh, R., Gong, J., & Jalili, N. (2022). Purposeful communication in human-robot collaboration: A review of modern approaches in manufacturing. *IEEE Access*, 10, 129344–129361.
- Sanfilippo, F., Azpiazu, J., Marafioti, G., Transeth, A. A., Stavdahl, Ø., & Liljebäck, P. (2017). Perception-driven obstacle-aided locomotion for snake robots: The state of the art, challenges and possibilities. *Applied Sciences*, 7(4), 336.
- Sanfilippo, F., Fossum, M., & Hansen, M. R. (2025). Redefining care and work practices: The role of autonomy in the future of nursing homes. *Proc. of the International Conference on eXtended Reality (XR Salento)*.
- Sanfilippo, F., Moosavi, S. K. R., & Zafar, M. H. (2025). Autonomy uprising: How civilians, first responders, and robots are redefining the future of search-and-rescue (sar). *Proc. of the 9th IEEE International Conference on Robotics and Automation Sciences (ICRAS)*, 1–8. <https://doi.org/10.1109/ICRAS65818.2025.11108899>
- Sanfilippo, F., Wiley, T., & Rousi, R. (2025). Robocup soccer autonomy uprising: How crowds, referees, and humanoid robots are redefining the future of human-robot interaction. *Proc. of the 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 1131–1139.
- Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human factors*, 58(3), 377–400.
- Sellner, B., Heger, F. W., Hiatt, L. M., Simmons, R., & Singh, S. (2006). Coordinated multiagent teams and sliding autonomy for large-scale assembly. *Proceedings of the IEEE*, 94(7), 1425–1444.
- Sigitas, K., Inga, L., Audrius, S., Marius, B., Mantas, M., Margarita, C., Jorge, R., Lilly, S., Varvara, L., Sophie, E., et al. (2024). Mood, cognitive function, and plasma kynurenine metabolites responses following severe changes in physical activity. *Medicine and Science in Sports and Exercise*.
- Tallam, K. (2025). Alignment, agency and autonomy in frontier ai: A systems engineering perspective. *arXiv preprint arXiv:2503.05748*.
- Tamminen, K. A., Palmateer, T. M., Denton, M., Sabiston, C., Crocker, P. R., Eys, M., & Smith, B. (2016). Exploring emotions as social phenomena among canadian varsity athletes. *Psychology of sport and exercise*, 27, 28–38.
- Vaassen, B. (2022). Ai, opacity, and personal autonomy. *Philosophy & Technology*, 35(4), 88.
- Verma, V., Maimone, M. W., Gaines, D. M., Francis, R., Estlin, T. A., Kuhn, S. R., Rabideau, G. R., Chien, S. A., McHenry, M. M., Graser, E. J., et al. (2023). Autonomous robotics is driving perseverance rover's progress on mars. *Science Robotics*, 8(80), eadi3099.
- Washington, R., Golden, K., Bresina, J., Smith, D. E., Anderson, C., & Smith, T. (1999). Autonomous rovers for mars exploration. *1999 IEEE Aerospace Conference. Proc. (Cat. No. 99TH8403)*, 1, 237–251.
- Wiese, E., Shaw, T., Lofaro, D., & Baldwin, C. (2017). Designing artificial agents as social companions. *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 1604–1608.
- You, S., Kim, J.-H., Lee, S., Kamat, V., & Robert Jr, L. P. (2018). Enhancing perceived safety in human-robot collaborative construction using immersive virtual environments. *Automation in construction*, 96, 161–170.
- You, S., & Robert, L. (2018a). Teaming up with robots: An imoi (inputs-mediators-outputs-inputs) framework of human-robot teamwork. *International Journal of Robotic Engineering*, 2(3).
- You, S., & Robert, L. (2018b). Trusting robots in teams: Examining the impacts of trusting robots on team performance and satisfaction. *Proc. of the 52th Hawaii International Conference on System Sciences*, Jan, 8–11.
- You, S., & Robert Jr, L. P. (2024). Trusting and working with robots: A relational demography theory of preference for robotic over human co-workers. *Management Information Systems Quarterly*, 48(4), 1297–1330.
- Zhang, Y., Ratnayake, T. S., Sew, C., Knibbe, J., Goncalves, J., & Johal, W. (2025). Can you pass that tool?: Implications of indirect speech in physical human-robot collaboration. *Proc. of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–18.