

Designing the Socio-Technical & Economic Protocols for Inclusive Off-World Settlement Commercial Training and Operational Frameworks for the Human-Robot Civilizational Shift

[author] -----
Veronica Chiaravalli, Founder, CEO, and CRO at Emerging Technologies Sweden,
vc.multiplanetaryx@gmail.com

[symposium, session] -----
Space Policy, Law, and Governance

[paper information] -----
☒ this paper was not published before

☐ I agree that Space Renaissance International will publish this paper on their websites and publications

ABSTRACT

Humanity is approaching two transformations simultaneously: the rapid development of artificial intelligence, autonomous agents and embodied robotics on Earth, and the expansion of sustained human and robotic activity beyond Earth. Their convergence raises a socio-technical question: how should emerging off-world settlements be designed when humans and increasingly capable intelligent machines must live, work and make decisions within the same operational systems?

This paper develops a conceptual framework centered on **autonomous pre-formation**: the possibility that robotic and autonomous systems establish infrastructure, workflows and operating assumptions before substantial human settlement, embedding technological choices before human requirements, governance and social arrangements are sufficiently integrated.

The paper proposes an **Off-World Transition Protocol (OWP)** supported by Systems Thinking, Futures Thinking and three analytical lenses: social-technological impact, exponential and converging technologies, and frontier/off-world environments. The framework combines **Human-Intelligent Systems Architecture (HISA)**, the **Athena Initiatives**, **Frontier Operational Integration Sprints**, adaptive reassessment of human-machine authority and capability, resilience when intelligent systems are degraded or unavailable, and economic participation.

The Moon is treated as an emerging socio-technological frontier where human performance, HRI, autonomy, resilience, governance and economic inclusion can be designed together before systems become entrenched.

Keywords: Human-Robot Interaction; Artificial Intelligence; Autonomous Systems; Human-Autonomy Teaming; Futures Thinking; Systems Thinking; Space Psychology; Lunar Economy; Off-World Settlement; Socio-Technical Systems

PAPER

1 Introduction: The Off-World Integration Challenge

Artificial intelligence is moving rapidly from information systems toward agents and embodied systems capable of increasingly complex action in digital and physical environments. At the same time, governments, research organizations and commercial actors are preparing for increasingly sustained human and robotic activity in cislunar space and on the lunar surface.

NASA's Moon to Mars Architecture illustrates the systems character of this transition. Its current architecture identifies interconnected sub-architectures including Autonomous Systems and Robotics, Human Systems, Habitation, Communications and Positioning, Navigation and Timing, Data Systems, Power, Logistics, Infrastructure Support and In-Situ Resource Utilization. Autonomous systems and robotics are explicitly intended both to assist crews and to operate during uncrewed periods.[1]

These developments should not be treated as unrelated technological trends. Together they may create an emerging socio-technological condition in which humans, artificial intelligence and embodied machines increasingly participate in the same operational, organizational and economic systems.

This requires a broader conception of technological design. NASA's human-integration guidance treats human-system integration as integral to crewed-system development,[2] while HRI is interdisciplinary, drawing upon robotics, AI, psychology, cognitive science, design and the social sciences.[3] Responsible Innovation adds anticipation, reflexivity, inclusion and responsiveness, emphasizing societal and ethical consequences while innovation remains capable of being shaped.[4]

This paper extends that principle into an off-world context.

Its central concern is the **Integration Gap**: a condition in which technological or autonomous capability develops faster than the human, organizational, leadership, governance and socio-economic capacity required to integrate it safely and meaningfully.

The problem becomes particularly consequential in frontier environments. Isolation, constrained resources, communication delays, limited evacuation possibilities, environmental hazards and dependence upon technological infrastructure can transform an integration failure that is inconvenient on Earth into a mission-critical event beyond Earth.

The Moon should therefore be understood not only as a technological frontier but as an emerging **socio-technological frontier**.

2 From Technological Acceleration to the Integration Gap

Technological innovation is frequently evaluated through capability, performance, efficiency, scalability and cost. These measures are necessary, but they are insufficient for systems in which humans and intelligent machines interact.

A system can succeed technically while generating undesirable consequences for human agency, privacy, trust, responsibility, psychological well-being, skills, employment or access. The relevant unit of analysis is therefore not simply the machine, but the **human-technological system surrounding the machine**.

This paper uses **Integration Gap** as an umbrella term for the condition in which technological capability advances faster than the surrounding human system can adapt.

The gap may appear when competencies lag behind system capability; organizational structures no longer reflect human-machine roles; responsibility becomes ambiguous; trust becomes miscalibrated; governance lags behind deployment; technological dependence increases; or economic participation becomes disconnected from technological productivity.

Human-autonomy research demonstrates the complexity of this relationship. O'Neill et al.'s review of empirical Human-Autonomy Teaming research examines humans and autonomous agents working interdependently toward shared goals and identifies factors including communication, training, team composition, task characteristics and characteristics of autonomous agents as important to effective teaming.[5]

Trust presents another integration challenge. Mehrotra et al. emphasize **appropriate trust** rather than simply maximizing trust in AI, examining the problem of calibrating human reliance on intelligent systems.[6]

These findings support a central proposition of this paper:

Increasing machine capability does not automatically produce effective human-machine integration.

2.1 Autonomous Pre-Formation

Autonomous pre-formation describes the possibility that robotic and autonomous systems may establish infrastructure, workflows, interfaces and operational assumptions before substantial human settlement.

Autonomous preparation itself is not the problem. Robotic systems may be essential for scouting, construction, resource utilization, inspection and maintenance while reducing human exposure to hazardous environments. NASA anticipates autonomous systems assisting crews and operating during uncrewed periods,[1][7] while its Moon to Mars objectives envision integrated human-robotic methods and interoperable lunar infrastructure supporting continuous robotic and human presence and a broader lunar economy.[8]

The concern, therefore, is not that robots may arrive first. The concern is that **technological and operational assumptions may become embedded in infrastructure before the requirements of the humans who will eventually inhabit those environments have been sufficiently incorporated**. This creates an important reversal in the historical relationship between humans and machines.

On Earth, increasingly capable intelligent machines are entering physical, organizational and economic environments historically designed and constructed by humans.

In emerging off-world environments, humans may instead enter environments that have already been substantially prepared, supported or maintained by intelligent machines.

This raises a socio-technical question that technological capability alone cannot answer:

If autonomous systems participate in constructing and operating an off-world settlement before substantial human habitation, when—and by whom—are the human requirements of that settlement defined?

Decisions concerning interfaces, authority, task allocation, accessibility, privacy, safety, psychological well-being, trust, economic participation and human performance may become embedded in infrastructure and operational practices long before a permanent human community exists to negotiate them.

HRI should therefore be designed around the locations and operational conditions in which humans and machines will eventually live and work together, rather than added after technological and automation architectures are already fixed.

Autonomous pre-formation consequently represents both a risk and an opportunity: the risk that technological architecture matures faster than human integration, and the opportunity to design human-machine relationships prospectively rather than retrofit them after systems become entrenched.

3 Method and Technological Context: Understanding an Accelerating Transition

The challenge addressed by this paper emerges not from artificial intelligence, robotics or automation considered separately, but from their accelerating **convergence, diffusion and embodiment**.

The methodological approach combines four perspectives:

Futures Thinking, used to identify signals, consider alternative futures and prepare for uncertainty;

Systems Thinking, used to examine interactions, feedback loops, dependencies, delays and unintended consequences;

Human-Robot Interaction and Human-Autonomy Teaming, used to examine relationships between humans and increasingly autonomous systems; and

Responsible Innovation, used to anticipate societal consequences while technological systems remain capable of being shaped.

Together, these perspectives move the analysis beyond the question of what a technology can do toward the broader question:

What happens to the human system when technological capability changes?

3.1 From Digital AI to Physical AI

Artificial intelligence has moved rapidly from specialist applications toward widely accessible systems capable of generating and analyzing information, supporting decisions and increasingly performing tasks through agentic systems.

A parallel transformation is occurring in robotics as machine intelligence increasingly becomes embodied in systems capable of perceiving, reasoning and acting within physical environments.

As intelligence acquires physical agency, questions of authority, communication, task allocation, trust, responsibility, verification and failure become more consequential.

The problem therefore increasingly becomes not simply:

What can the machine do?

but:

How should humans and increasingly capable intelligent machines live, work, decide and adapt together?

3.2 Adoption Itself Is Changing

The diffusion of generative AI provides an important contemporary signal. Traditional technology-adoption frameworks, including Rogers' diffusion model and Moore's *Crossing the Chasm*, remain useful for understanding movement from innovators and early adopters toward broader markets.[9][10] Intelligent technologies, however, introduce an additional dynamic. **AI is not only an innovation being adopted. AI is increasingly becoming part of the infrastructure through which other innovations are researched, created, communicated, marketed, sold and supported.** AI increasingly contributes to research, product development, content production, market analysis, customer interaction and commercialization. Marketing and sales therefore become part of the acceleration mechanism.

The relationship can be conceptualized as:

AI capability → faster innovation and production → AI-supported communication and commercialization → faster diffusion → greater adoption and investment → further technological development.

The strength and speed of this feedback relationship will differ between technologies and industries and should not be interpreted as an inevitable exponential trajectory. Its significance is that **the mechanisms through which technological innovation diffuses are themselves becoming technologically augmented.** This may require innovation-adoption theory to increasingly distinguish between the speed at which people gain access to a technology and the much deeper process through which organizations and societies successfully integrate it.

3.3 From the Innovation Chasm to Recurring Integration Thresholds

Rapid adoption does not imply successful integration. Individual use of generative AI may begin quickly, while organizational integration requires workflows, accountability, skills, evaluation, security and governance. Autonomous agents add questions of delegated authority and oversight; embodied AI adds physical safety, maintenance and HRI; and off-world deployment adds isolation, communication delay, constrained resources and potentially life-critical technological dependence. Technological transitions may therefore produce **recurring integration thresholds**:

Digital AI assistance → organizational AI integration → agentic autonomy → embodied/physical AI → human-machine socio-economic integration.

Each transition introduces different human and organizational requirements. The relevant question consequently becomes not only:

Has the technology crossed the adoption chasm?

but:

Can human, organizational and social systems repeatedly adapt as technological capability crosses successive thresholds?

This distinction defines the Integration Gap more precisely. The Integration Gap is not simply resistance to innovation. It is the **asymmetry between the rate at which technological capability and diffusion change and the rate at which human competencies, organizational structures, governance, resilience and social arrangements can adapt.**

3.4 High-Disruption Scenarios and the Intelligence Inversion

The potential economic and social consequences of this acceleration are substantial but uncertain. Emad Mostaque develops what he describes as an **Intelligence Inversion**: a scenario in which intelligence moves from being relatively scarce and expensive toward becoming increasingly abundant and inexpensive through AI. His work *The Last Economy* explores the consequences this could have for economic systems historically structured around scarce human cognitive labor.[11] Mostaque has extended this argument in public interviews into high-acceleration scenarios involving autonomous agents, occupational displacement, inequality and economic disruption.[12] This paper does **not** treat such projections as predictions. They are used as a **high-disruption scenario** against which existing systems can be stress-tested.

Contrasting futures remain plausible: AI may complement rather than replace many forms of work; adoption may remain uneven; robotics may encounter technical, economic, regulatory and social constraints; and new occupations may emerge. Futures Thinking therefore avoids assuming that any single scenario represents the future.

The relevant principle is:

Uncertainty does not eliminate the need for preparation.

3.5 Futures Thinking, Systems Thinking and the Exponential Mindset

This paper uses Futures Thinking rather than technological prediction. Approaches taught by Jane McGonigal through the Institute for the Future emphasize signals of change, alternative possibilities, simulation and action-oriented foresight.[13] Systems Thinking complements this approach by examining feedback, dependencies, delays and relationships generating system behavior.[14] Applied to AI and robotics, it directs attention toward interactions among technology, human capability, organizations, markets, infrastructure, regulation, education and leadership. The Exponential Organizations framework adds the organizational challenge of adapting to technologies capable of unusually rapid scaling.[15] Here, an exponential mindset means not pursuing speed indiscriminately, but recognizing that organizations designed around slower technological cycles may require substantially greater adaptability.

In high-stakes environments: **speed must be coupled with foresight, resilience, accountability and human performance.**

3.6 The Three Analytical Lenses

These methodological foundations are operationalized through three interacting analytical lenses.

Lens 1 — Social and Technological Impact examines capability alongside consequences for agency, privacy, leadership, responsibility, psychological well-being, employment, economic participation and required

human competencies. It treats technological and social innovation as components of the same evolving system rather than addressing social consequences only after deployment.

Lens 2 — Exponential and Converging Technologies examines interactions among AI, robotics, autonomous agents, communications, immersive environments, digital twins, advanced simulation and related technologies. Their convergence may produce capabilities difficult to anticipate from any technology independently; meaningful changes in capability should therefore trigger reassessment of training, authority, responsibility and human-machine interaction.

Lens 3 — Frontier and Off-World Environments examines these developments under isolation, constrained resources, communication delays, environmental hazards, limited evacuation and technological dependence. NASA's Human Factors and Behavioral Performance research addresses several of these risks, including isolation, cognition, team performance, training and human-robotic interactions.[16]

These conditions amplify the consequences of integration failures.

The greatest analytical value therefore lies at the intersection of the three lenses:

accelerating and converging technological capability + social and economic consequences + frontier-environment constraints.

It is at this intersection that the Integration Gap becomes particularly consequential.

4 The Off-World Transition Protocol

The **Off-World Transition Protocol (OWP)** is an adaptive socio-technical framework for preparing organizations and crews for increasingly autonomy-integrated off-world operations. Rather than treating governance as static, it combines stable principles with operational protocols that can be reassessed as machine capability, mission conditions and social impacts change.

A simplified adaptive cycle is:

Technological or operational change → human-impact assessment → authority and responsibility reassessment → competency reassessment → training and simulation → protocol update → validation → deployment → monitoring → reassessment.

Significant changes in machine capability become reassessment triggers. A robot whose hardware remains unchanged may acquire materially different capabilities through software, AI models, sensing or autonomy updates; human roles, training and authority structures should therefore be reassessed when those capabilities change.

The protocol also incorporates **economic inclusion**. NASA's Moon to Mars objectives explicitly envision lunar infrastructure in which industry and international partners can maintain continuous robotic and human presence and participate in a lunar economy in which NASA is not the sole user.[8] This emerging economy raises questions extending beyond whether robots can excavate, inspect, construct or maintain infrastructure:

Who owns productive systems? Who controls them? Who can access their capabilities? Who participates in the value created?

Smaller companies, universities, researchers, entrepreneurs and remote participants should not necessarily require ownership of entire robotic fleets or transportation architectures to participate.

Service-based, shared-access and distributed operational models therefore warrant investigation as mechanisms for broadening participation in an off-world economy.

5 Human-Intelligent Systems Architecture

Human-Intelligent Systems Architecture (HISA) is the proposed organizational and operational layer of the framework.

Its central question is:

Who decides what as autonomy scales?

HISA focuses on decision authority, responsibility attribution and traceability, human-in/on-the-loop governance, override and escalation structures, trust calibration, capability awareness and failure-mode anticipation.

NASA's Human Integration Design Handbook and Human Integration Design Processes reinforce the broader principle that humans and systems must be successfully integrated during development of crewed systems and operations.[2]

Human-autonomy teaming research similarly demonstrates the importance of interdependence, communication and training.[5]

HISA therefore treats authority as something that must be **designed deliberately**, rather than inferred automatically from technological capability.

Trust is similarly treated as a dynamic operational variable. The objective should not be maximum trust in AI, but **appropriate reliance**.

A practical HISA implementation could therefore include a **trust-calibration log** documenting where reliance upon intelligent systems is appropriate, where verification is required, where humans must retain authority and how trust should be reassessed following anomalous behavior, failure or significant capability change.

HISA also recognizes a **capability-synchronization problem**.

If machine capability changes faster than operators and leaders can update their understanding, training, procedures and authority structures, the Integration Gap can become safety-critical.

HISA must consequently account for movement in both directions: increasing delegation to autonomous systems when capability and circumstances justify it, and safe reassumption of functions by humans when automation becomes degraded, unavailable or inappropriate.[22]

6 Commercial Training for the Human-Robot Transition

Technological architecture alone cannot resolve the Integration Gap.

Organizations and crews require training that develops the human capabilities necessary to work effectively with increasingly autonomous systems while retaining sufficient competence to operate when those systems are degraded or unavailable.

The **Athena Initiatives** provide the commercial training layer of the proposed framework.

Athena integrates four domains:

Human Performance in Space; HRI Integration; Autonomy Acceleration Literacy; and Operating Without AI.

6.1 Human Performance in Space

Human Performance in Space addresses the biological, psychological and interpersonal realities of operating in isolated, confined and high-risk environments. Kanas' work on the psychological challenges of spaceflight examines psychosocial stressors, psychological and interpersonal issues, cultural factors, countermeasures, crew-ground interaction and increasing autonomy during longer-duration missions.[17] NASA's Human Research Program similarly investigates isolation, cognition, team cohesion, training and psychological readiness for exploration missions.[16]

These factors cannot be separated from HRI. A robot that performs optimally from an engineering perspective may still increase cognitive workload, create inappropriate trust, disrupt team coordination or generate uncertainty about responsibility. Conversely, properly designed intelligent systems may reduce workload, improve situational awareness and support crews where human resources are constrained.

6.2 HRI Integration

HRI Integration develops competencies for task allocation, communication, trust, authority, verification and coordination with embodied and autonomous systems. Because HRI is inherently interdisciplinary,[3] training should not be limited to robotic operation. Leaders and teams require sufficient understanding of robotic capability and limitations to determine when automation should act independently, when humans should supervise, when intervention is required, how machine outputs should be verified and how responsibility is distributed when humans and autonomous systems jointly produce outcomes. Leadership in an AI- and robotics-intensive environment increasingly requires the ability to understand not only human teams but **human-machine teams**.

6.3 Autonomy Acceleration Literacy

Autonomy Acceleration Literacy prepares participants to recognize that machine capability and organizational assumptions may change at different speeds. A system deployed today may not possess the same operational capabilities after successive model, software or autonomy updates. Training should therefore develop the ability to recognize signals of capability change, reassess assumptions, identify new failure modes, understand changing human-machine boundaries and determine when technological changes require corresponding changes in procedures, competencies, authority or governance. This is where Futures Thinking and exponential mindset become operational rather than theoretical. The objective is not to predict exactly what AI or robotics will be capable of. It is to create organizations capable of recognizing when **reality has begun to diverge from the assumptions under which their systems were designed**.

6.4 Operating Without AI

Operating Without AI addresses two related forms of resilience: resilience of the technological system and resilience of the human beings operating within it. AI-enabled operations depend upon an interconnected technical stack that may include electrical power, communications, computing resources, software, networks, sensors, data and remote or cloud-based services. War, cyber incidents, infrastructure damage, severe natural events, space weather, equipment failure or other disruptions can compromise one or more of these dependencies. Extreme disruption scenarios, including electromagnetic pulse effects, further illustrate why survival-critical operations should not assume uninterrupted access to advanced digital capability. The objective is not to predict which disruption will occur. It is to recognize that:

high dependence combined with inadequate fallback capability creates systemic vulnerability.

The second risk is human capability degradation. Cognitive offloading—the transfer of cognitive work to external tools—can improve immediate performance, but research also identifies potential costs when access to the external resource is unexpectedly lost.[18][19][23][24] Recent research examining AI-assisted cognition goes further: evidence suggests that offloading cognition to AI can impede skill acquisition or contribute to skill decay depending on how the technology is used.[20] This creates a particularly consequential off-world risk. Consider a future lunar or planetary settlement in which AI and autonomous systems routinely support diagnostics, resource allocation, environmental monitoring, maintenance, navigation, scientific analysis and operational decision-making. If those systems become unavailable, technological redundancy alone may be insufficient if humans have simultaneously lost practice in the competencies required to assume essential functions.

The critical failure condition is therefore not simply:

Loss of AI capability

but:

Loss of AI capability + insufficient retained human capability.

In an isolated off-world environment, crews must also confront questions that are easily overlooked in terrestrial technology discussions: Is the necessary expertise physically present? Is critical knowledge locally accessible if communication with Earth is unavailable? Can another crew member assume the function if the specialist is incapacitated? How long must the settlement remain operational before external assistance, replacement equipment or physical resupply can arrive? NASA's current human-system standard explicitly applies to space vehicles and habitats, including lunar and planetary habitation, and establishes requirements intended to minimize health and performance risks to crews.[21]

Operating Without AI therefore requires **resilient interdependence** rather than technological rejection: using advanced intelligent systems extensively while deliberately preserving essential human competence, locally available knowledge, cross-training, degraded-mode procedures and appropriate technological redundancy. Future high-stakes teams consequently require dual readiness:

the ability to operate effectively with advanced AI and autonomous systems, and the ability to remain operationally competent when those systems cannot be used.

7 Frontier Operational Integration Sprints

Frontier Operational Integration Sprints provide the implementation and testing mechanism for the framework. These are short-cycle operational exercises through which teams prototype and test resource-aware human-machine workflows under realistic constraints. They differ from conventional design sprints because their purpose is not merely to accelerate product development. Their purpose is to expose **integration failures before deployment** and generate evidence for revising HISA, Athena training and the Off-World Transition Protocol. Sprints can be conducted through simulation, immersive environments, digital twins, robotics simulation, terrestrial analog environments and eventually operational robotic systems. Possible scenarios include loss of AI assistance, unexpected increases or decreases in autonomy, communications delay, power constraints, conflicting human and machine recommendations, sensor degradation, human cognitive overload, robotic failure and emergency transfer of authority.

Each sprint can test operational questions:

Who has authority? Who verifies the system? Who intervenes? How is trust recalibrated after an error? What happens when AI assistance disappears? Can humans still perform mission-critical functions? What happens when a machine acquires capabilities that were absent when the original procedures were written?

Results feed back into HISA, Athena and the Off-World Transition Protocol.

8 Economic Inclusion and Meaningful Human Participation

Technological acceleration raises a further question that cannot be resolved through HRI alone: **How will humans participate economically in systems in which intelligent machines perform an increasing share of productive activity?** Near-term responses may include new professional pathways, HRI occupations, remote supervision, robotics operations, entrepreneurship and training. These are important, but this paper treats them as **stage-dependent rather than permanent solutions**. If machine capability continues increasing, creating human occupations around temporary technological limitations may not provide durable economic participation. Mostaque's Intelligence Inversion scenario is useful because it forces examination of what happens if machine intelligence becomes substantially more abundant.[11]

Meaningful participation can include employment, but it can also include entrepreneurship, ownership, investment, governance, education, scientific participation, creative production and access to productive technological systems. This distinction becomes particularly important in an off-world economy. If productive lunar infrastructure requires expensive transportation, computing, communications and robotic fleets, participation could become concentrated among a relatively small number of actors. NASA's objective of developing lunar infrastructure involving industry and international partners beyond NASA as sole user provides an important institutional starting point.[8]

The next question is:

How can that economy develop while enabling meaningful participation by smaller organizations and individuals?

One possible direction is **access rather than ownership**. Organizations might purchase robotic services, operational time, remote experimentation, simulation capacity, data services or access to shared infrastructure without owning the complete technological stack. This paper does not attempt to resolve wider terrestrial debates over universal basic income, robot taxation or post-labor economics.

The narrower proposition is:

Economic architecture should be considered alongside technological and operational architecture from the early stages of off-world settlement design.

9 Discussion: Toward Adaptive Pioneering Innovation

The framework developed here treats technological innovation and social innovation as mutually dependent. Technological innovation without social understanding can optimize systems while overlooking agency, privacy, psychological well-being, responsibility, resilience or economic access. Social innovation without sufficient technological understanding can design interventions for technological conditions that may already be changing. Futures Thinking without implementation can remain speculative. Speed without Systems Thinking can amplify unintended consequences. Automation without human-performance design can create dependency. And resilience without technological innovation can prevent organizations from taking advantage of capabilities that may substantially improve safety and performance.

The challenge is therefore not to choose between technological acceleration and human-centered development. It is to increase humanity's capacity to **anticipate, understand, integrate, adapt, retain essential capability and participate meaningfully in the value created by technological systems**. The emerging off-world economy provides an unusual opportunity to apply this principle prospectively.

Instead of:

Technology → Deployment → Social consequences → Delayed response

the proposed approach becomes:

Signals → Foresight → Socio-technical analysis → Integration Gap diagnosis → Co-design → Training → Simulation → Deployment → Monitoring → Reassessment → Adaptation

This paper uses **pioneering innovation** provisionally to describe action at the boundary between present systems and emerging conditions: identifying assumptions likely to fail beyond the present frontier and beginning to design, test and adapt healthier socio-technological systems before new arrangements become entrenched.

10 Conclusion

Artificial intelligence, autonomous agents and embodied robotics may transform organizations, work and economic participation during the same historical period in which humanity begins constructing increasingly sustained activity beyond Earth. The final configuration of either transition cannot yet be known. That uncertainty argues against both technological determinism and static planning.

The central proposition of this paper is that **emerging off-world systems should be designed as socio-technical systems from the beginning**. Frontier environments amplify the consequences of poor integration, while autonomous pre-formation creates the possibility that technological systems will establish infrastructure and operational assumptions before substantial human communities arrive. The response proposed here combines an adaptive **Off-World Transition Protocol**, **Human-Intelligent Systems Architecture**, commercial training through **Athena**, operational experimentation through **Frontier Operational Integration Sprints**, resilient interdependence through **Operating Without AI**, and explicit attention to **economic inclusion**.^[25] The goal is neither maximum automation nor maximum human involvement. It is a healthy human-technology ecosystem capable of adapting as technology, missions and social conditions change. Earth is increasingly integrating intelligent machines into environments historically constructed around humans.

The Moon may present the inverse challenge:

humans entering environments substantially prepared and maintained by intelligent machines.

That inversion creates a rare design opportunity. Humanity does not have to reproduce terrestrial socio-technological arrangements automatically beyond Earth. It can begin asking, before those systems become entrenched:

What kind of human-machine civilization do we want to build?

ACRONYMS AND KEY TERMS

Acronym	Description
AI	Artificial Intelligence
HAT	Human-Autonomy Teaming
HRI	Human-Robot Interaction
HISA	Human-Intelligent Systems Architecture
OWP	Off-World Transition Protocol

Integration Gap — the asymmetry between the rate of technological capability and diffusion and the ability of human, organizational, governance and socio-economic systems to adapt.

Autonomous Pre-Formation — the establishment of infrastructure, workflows or operational assumptions by autonomous systems before substantial human habitation.

Autonomy Acceleration Literacy — the ability to recognize and respond to changes in machine capability that may invalidate existing operational assumptions.

Operating Without AI — preservation of human and organizational capability when intelligent assistance becomes degraded or unavailable, while maintaining effective human-AI collaboration during normal operations.

Resilient Interdependence — a human-machine relationship that uses advanced technological capability while preserving sufficient human, knowledge and technological redundancy to maintain critical operations during degradation or failure.

REFERENCES

- [1] National Aeronautics and Space Administration. (2026). *Moon to Mars Architecture: Components*. NASA.
- [2] National Aeronautics and Space Administration. (2026). *Human Integration Design Handbook (HIDH), Revision 1*. Office of the Chief Health and Medical Officer, NASA.
- [3] Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., & Šabanović, S. (2024). *Human-Robot Interaction: An Introduction* (2nd ed.). Cambridge University Press.
- [4] Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580.
- [5] O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2022). Human-autonomy teaming: A review and analysis of the empirical literature. *Human Factors*, 64(5), 904–938. <https://doi.org/10.1177/0018720820960865>.
- [6] Mehrotra, S., Degachi, C., Vereschak, O., Jonker, C. M., & Tielman, M. L. (2024). A systematic review on fostering appropriate trust in human-AI interaction: Trends, opportunities and challenges. *ACM Journal on Responsible Computing*, 1(4), Article 26. <https://doi.org/10.1145/3696449>.
- [7] National Aeronautics and Space Administration. (2026). *Planning, Architecture & Analysis: Surface Robotics and Mobility*. Johnson Space Center.
- [8] National Aeronautics and Space Administration. (2026). *Moon to Mars Architecture: Strategy and Objectives*. NASA.
- [9] Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). Free Press.
- [10] Moore, G. A. (2014). *Crossing the Chasm: Marketing and Selling Disruptive Products to Mainstream Customers* (3rd ed.). HarperBusiness.
- [11] Mostaque, E. *The Last Economy: A Guide to the Age of Intelligent Economics*. Intelligent Internet.
- [12] Mostaque, E. (2025). Interview with Myriam François on artificial intelligence, automation and economic transformation. *The Tea with Myriam François*.
- [13] McGonigal, J., & Institute for the Future. *Futures Thinking Specialization*. Coursera.
- [14] Meadows, D. H. (2008). *Thinking in Systems: A Primer*. Chelsea Green Publishing.

- [15] Ismail, S., Malone, M. S., & van Geest, Y. (2014). *Exponential Organizations: Why New Organizations Are Ten Times Better, Faster, and Cheaper Than Yours (and What to Do About It)*. Diversion Books.
- [16] National Aeronautics and Space Administration. (2025). *Human Factors and Behavioral Performance*. Human Research Program, NASA.
- [17] Kanas, N. (2015). *Humans in Space: The Psychological Hurdles*. Springer.
- [18] Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688. <https://doi.org/10.1016/j.tics.2016.07.002>.
- [19] Richmond, L. L., & Taylor, R. G. (2025). The benefits and potential costs of cognitive offloading for retrospective information. *Nature Reviews Psychology*, 4, 312–321.
- [20] Cash, T. N., Kelly, M. O., Macnamara, B. N., & Risko, E. F. (2026). Is AI making us stupid? *Trends in Cognitive Sciences*, 30(8), 673–675. <https://doi.org/10.1016/j.tics.2026.06.004>.
- [21] National Aeronautics and Space Administration. (2026). *NASA-STD-3001 Volume 2: Human Factors, Habitability, and Environmental Health*. NASA.
- [22] Hauptman, A. I., Flathmann, C., & McNeese, N. J. (2024). Adapting to the human: A systematic review of a decade of human factors research on adaptive autonomy. *Applied Ergonomics*, 120, 104336.
- [23] Burnett, L. K., & Richmond, L. L. (2026). Meta-analytic investigations of the effect of cognitive offloading on memory-based task performance and interindividual variability. *Memory & Cognition*, 54, 144–168. <https://doi.org/10.3758/s13421-025-01743-8>.
- [24] Skulmowski, A. (2023). The cognitive architecture of digital externalization. *Educational Psychology Review*, 35, Article 101.
- [25] Chiaravalli, V. / Emerging Technologies Sweden. *Emerging Technologies Sweden / MultiplanetaryX*. Author-developed frameworks associated with the Integration Gap, HISA, Athena Initiatives, Frontier Operational Integration Sprints and off-world socio-technical experimentation.