Emerging Autonomy Solutions for Human and Robotic Deep Space Exploration

MATTHEW B. LUEBBERS*, CHRISTINE T. CHANG*, and AAQUIB TABREZ*, Department of Computer Science, University of Colorado Boulder, USA
JORDAN DIXON*, Smead Aerospace Engineering Sciences, University of Colorado Boulder, USA
BRADLEY HAYES, Department of Computer Science, University of Colorado Boulder, USA

The space community has traditionally been extremely cautious about the usage of autonomy for high stakes applications. Autonomy deployed in control systems has often been deterministic and verifiable, contrasting with modern, non-deterministic learning or interaction-based techniques. This is justifiable, as the cost of mission failure is exceptionally high. On the other hand, deterministic models inherently restrict the designs of deep space missions and preclude many exciting concepts. In this work, we briefly describe the challenges faced in deploying autonomy in deep space, and present two case studies to showcase where more advanced models or human-robot techniques would be useful. We also discuss a collection of emerging technologies that could be leveraged to hedge the risks of autonomy deployment for future deep space applications.

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Dependable and fault-tolerant systems and networks; Reliability; • **Human-centered computing** → **HCI theory, concepts and models**; **Models of learning**; • **Computing methodologies** → *Model development and analysis.*

Additional Key Words and Phrases: space autonomy, human-robot interaction, explainable AI, adaptive/adaptable autonomy

ACM Reference Format:
Matthew B. Luebbers, Christine T. Chang, Aaquib Tabrez, Jordan Dixon, and Bradley Hayes. 2021. Emerging Autonomy Solutions for Human and Robotic Deep Space Exploration. In *Proceedings of Workshop on SpaceCHI : Human-Computer Interaction for Space Exploration (SpaceCHI '21).* ACM, New York, NY, USA, 5 pages.

1 MOTIVATION

Interplanetary exploration and space robotics pose a set of unique challenges related to resource constraints and sustainability requirements. Efficiently combining the capabilities of astronauts, remote operators, and robotic assets into human-machine teams is a must. However, applications of human-robot interaction (HRI) for autonomous and semi-autonomous space operation have ample open questions. Solar system exploration will continue to require human interaction and oversight, encompassing a continuum of human-robot spatio-temporal relationships depending on mission type. As the complexity of this operating theater evolves, it is essential that humans trust and understand the systems with which they collaborate, and that systems can accommodate rapidly changing needs. We will discuss emerging technologies in HRI that can substantially contribute to these capabilities in the future.

2 CASE STUDIES

We begin by presenting two case studies, the first representing real-time, co-located teaming between astronauts and robotic assets in the presence of substantial time delays with mission control, and the second representing a deep space, remote interaction, involving uncrewed robotic assets with more infrequent HRI capacity. Both case studies highlight different autonomy challenges within the continuum of spatiotemporal HRI relationships.

*These authors contributed equally to this research.

2.1 Real-Time, Co-located: Mars

At least since Fitt's (1951) proposed "Men-are-better-at [sic], Machines-are-better-at (MABA-MABA)" [8], it has been accepted that a well-designed symbiotic relationship between human and machine can take advantage of their differing strengths and capabilities. Picture a future human expedition to a remote planetary body, such as the Martian surface, which has yet only been explored by teleoperated semi-autonomous vehicles with substantial time lags. With so many needs and great uncertainty, such a mission stands to benefit greatly from automated robotic assets. They can assist with construction, infrastructure support, science exploration, in-situ resource utilization, and emergency medical capabilities [9], making the resource-constrained mission more efficient and safe. A well-rounded team may consist of multiple astronauts and a variety of robotic assets with complementary capabilities. To this end, the agent interactions need to be carefully and flexibly designed in order for the automation to contribute positively to the mission.

The HRI design space continues to evolve as humans and machines collaborate more closely in shared-environments. For example, human and robot teammates may benefit from performing tasks together during habitat construction such as carrying heavy or awkwardly shaped equipment [17]. A few of the most fundamental HRI challenges facing this type of collaboration are succinctly described as including bi-directional intent recognition, dynamic task allocation, cost-function/task optimality adjustments based on co-worker preference or situational demands, and self-evaluation of automated agents [11]. Additionally, with respect to safety, an accurate and shared mental model [24] of teammate spatial proximity is increasingly important in the extreme spaceflight environment where astronauts must wear soft, pressurized spacesuits with limited mobility and fields of view. A robotic asset assisting an astronaut risks injuring the astronaut or potentially damaging the spacesuit. Mission critical risk such as in this example of a seemingly trivial task will hinder deployment of symbiotic human-robotic teams for future space exploration.

2.2 Deep Space, Remote: Europa

There is significant heritage in robotic exploration of the outer solar system, beginning with Pioneer 10's flyby of Jupiter in 1973. No spacecraft, however, has engaged in significant surface exploration further than Mars. The only probe to land on an outer solar system celestial body was Huygens on Titan in 2005, though the lander was immobile, and only successfully sent data for slightly over two hours on the surface. These mission concepts (flyby, orbital, short-term stationary lander) involve relatively low amounts of plan uncertainty, and therefore can rely on rigid, long-term command sequences planned weeks or months in advance. This advance planning protocol fulfills outer solar system operations requirements such as hours-long round-trip communication time, limited data transfer rate, and lack of significant orbital assets with which to relay signals, leading to communications blackouts.

Consider a mobile robot on Europa, tasked with exploring specific features over a period of weeks. The robot's functional time on the surface is limited by the intense radiation environment surrounding Jupiter [7]. This factor means the robot is solely battery powered, its designers opting not to spend weight on power generation hardware, placing a hard limit on the potential activities of the robot. Each communication event is therefore tremendously valuable and costly, both in terms of time and power. Uncertainty is too high to pre-program the robot's traverses, but the communication constraints preclude a high level of operator involvement akin to current Mars rover-like operation [19]. The mission must therefore deploy flexible autonomy capable of operating for long stretches without human intervention, both in terms of fault prevention/recovery and goal selection. It is also highly desirable that this autonomy is effectively explainable to its human operators, to allow for more trusted high-level planning and quicker response in failure scenarios.

3 EMERGING TECHNOLOGIES FOR FUTURE SPACE AUTONOMY

This section briefly describes the emerging technologies in machine learning and robotics that can be leveraged to hedge the risks described above encountered when employing autonomous systems in the deep space environment.

3.1 Explainable AI

Explainable AI (xAI) focuses on methods and techniques that facilitate human-interpretability of complex models as opposed to traditional "black box" models [10]. xAI methods allow modern machine learning algorithms to be responsibly used for high stakes applications (e.g., autonomous driving, robotics, medical diagnosis) by enabling users to better understand their decision-making [6, 18]. Thus, xAI methods that provide certain assurance and trust have started to gain more acceptance and popularity within the human-autonomy community, and therefore, we believe that they will play a crucial role in modern space autonomy [14].

Assessing complex models to determine trust levels prior to deployment is one widespread application of xAI. State-of-the-art approaches like FactSheets [1] or model cards [20] provide a quick reference to users describing the system's performance overview, including information about any model's provenance, details about its sensing and perception mechanisms, and intended operational domains. For example, surface robot explorers can be pre-trained in a simulated Europa environment to have detailed information on how they would behave in a specific scenario, enabling engineers to make informed decisions during deployment. Another popular application of xAI is mitigating failures and providing alternative action as recourse [3, 26]. One favored xAI technique is counterfactual explanation, which, given an input and target output, indicates how the input must change to alter the current output to the target output [5]. Another popular technique is providing semantic justification (giving the user the underlying decision-making rationale) for any reported observations and suggested plan contingencies for establishing trust in the system [12, 26]. These techniques are particularly informative in failure recovery scenarios, such as when a rover's initially planned traverse is rendered unfeasible. Furthermore, xAI has emerged as a crucial element in HRI for improving fluency and teamwork in human-robot collaboration scenarios while establishing a shared mental model [28]. For example, in control handover situations (e.g., where a semi-autonomous teleoperated system transfers control to a human operator due to its inability to operate in uncertain conditions), an explainable system can provide additional situational knowledge to its operators enabling informed and fluent handover of control [6, 27].

3.2 Virtual, Augmented, and Mixed Reality

Virtual, augmented, and mixed reality (VAMR) will be key components of HRI and autonomy as we head into the future of space exploration. While prior work applying VAMR to HRI is generally done in close proximity and without a significant time delay [13, 22, 23, 29, 31], with appropriate focus, VAMR can be adequately developed to assist both proximal and remote HRI. VAMR currently provides nearby operators with supplemental information and visual aids, increased situational awareness, and additional functionality and modes of communication [16, 23, 30]. These same tools can be applied for operational uses in both a collaborative Martian task or a more remote task on Europa. During a Mars EVA, a human-robot team might be exploring an area together. The human, aided by navigation cues and object recognition in their AR heads-up display, can guide the rover or quadcopter to conduct further investigation. In a cooperative construction task such as the one described above, information can be displayed in AR to provide insight into the robot's decisions to improve situational awareness, and visual aids might provide a shared mental model. Alternately, a remote operator planning a traverse on Europa could be experiencing the setting of the robot via VR, enabling them to program and execute commands using waypoints placed from a first person or overhead perspective. Planned paths can be previewed in VR with potential issues highlighted in the user's field of view. This experience also increases the operator's situational awareness of the robot's environment. Meanwhile, repeated use and training can allow the human to understand the autonomy decisions and capabilities from the visual perspective of the robot. In both situations, the use of VAMR provides additional safety for both the humans and robots.

3.3 Adaptive & Adaptable Automation

Adaptive control refers to a computer, in this case a robotic teammate, automatically adjusting controller parameters as a function of some measurement(s). Adaptable control, in contrast, refers to system changes made manually by the supervisory (human) operator [24]. These are not mutually exclusive design concepts, and useful taxonomies have been proposed involving characterizations of 'levels of automation' (LOA) [21]. They have served as useful starting places for designing an optimal balance between the utility of self-adjusting, adaptive robotic teammates, and the agency maintained by the human teammates through adaptable capabilities.

A wide variety of research has been conducted exploring the potential of adaptive and adaptable systems to enhance operator situational awareness, reduce mental workload, and improve efficiency of operations across different domains [2, 15]. Unfortunately, such systems are also liable to cause detrimental effects on pilot and operator performance and trust with various implementation designs [4, 21]. Conceptually, a well-designed adaptive-adaptable robotic teammate could contribute to effective dynamic task allocation, make intelligent adjustments to task optimality standards to best integrate with their human counterparts, and ideally detect system failures through self-evaluation. For all the research performed over the recent decades, we have just began scratching the surface on how to design increasingly complex automation frameworks that function in increasingly uncertain (shared-)environments. These limitations are concisely summarized by Sheridan [25] who is credited with originating the idea of LOA. While Sheridan suggests we may never reach a point in engineering design where "experience, art, and iterative trial and error" ([25], p.27) are not involved in the automation development process, there remains vast potential for facilitating safety and efficiency in future space exploration missions.

4 CONCLUSION

This work has provided a brief overview of problems faced in the deployment of autonomy for deep space missions, grounded in two case studies – one describing challenges in the inner solar system and another for the outer solar system. We then highlighted key emerging technologies that can be leveraged to alleviate these challenges: explainable AI; virtual, augmented, and mixed reality; and adaptive & adaptable automation. Utilizing these technologies can make deep space autonomy more feasible, thus increasing mission capabilities across the solar system.

REFERENCES

[1] Matthew Arnold, Rachel KE Bellamy, Michael Hind, Stephanie Houde, Sameep Mehta, Aleksandra Mojsilović, Ravi Nair, K Natesan Ramamurthy, Alexandra Olteanu, David Piorkowski, et al. 2019. FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development* 63, 4/5 (2019), 6–1.

[2] Nathan R Bailey, Mark W Scerbo, Frederick G Freeman, Peter J Mikulka, and Lorissa A Scott. 2006. Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human factors* 48, 4 (2006), 693–709.

[3] Tathagata Chakraborti, Sarath Sreedharan, Yu Zhang, and Subbarao Kambhampati. 2017. Plan explanations as model reconciliation: Moving beyond explanation as soliloquy. *arXiv preprint arXiv:1701.08317* (2017).

[4] Ewart de Visser and Raja Parasuraman. 2011. Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making* 5, 2 (2011), 209–231.

[5] Finale Doshi-Velez, Mason Kortz, Ryan Budish, Chris Bavitz, Sam Gershman, David O'Brien, Kate Scott, Stuart Schieber, James Waldo, David Weinberger, et al. 2017. Accountability of AI under the law: The role of explanation. *arXiv preprint arXiv:1711.01134* (2017).

[6] Na Du, Jacob Haspiel, Qiaoning Zhang, Dawn Tilbury, Anuj K Pradhan, X Jessie Yang, and Lionel P Robert Jr. 2019. Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload. *Transportation research part C: emerging technologies* 104 (2019), 428–442.

[7] Paul D Fieseler, Shadan M Ardalan, and AR Frederickson. 2002. The radiation effects on Galileo spacecraft systems at Jupiter. *IEEE Transactions on Nuclear Science* 49, 6 (2002), 2739–2758.

[8] Paul M. Fitts. 1951. Human engineering for an effective air-navigation and traffic-control system. (1951).

[9] Terrence Fong and Illah Nourbakhsh. 2005. Interaction challenges in human-robot space exploration. *Interactions* 12, 2 (2005), 42–45.

[10] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web* 2, 2 (2017).

[11] Bradley Hayes and Brian Scassellati. 2013. Challenges in shared-environment human-robot collaboration. *Workshop on Collaborative Manipulation at the International Conference on Human-Robot Interaction* (2013).

[12] Bradley Hayes and Julie A Shah. 2017. Improving robot controller transparency through autonomous policy explanation. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. IEEE, 303–312.

[13] Hooman Hedayati, Michael Walker, and Daniel Szafir. 2018. Improving Collocated Robot Teleoperation with Augmented Reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. Association for Computing Machinery, Chicago, IL, USA, 78–86. https://doi.org/10.1145/3171221.3171251

[14] Brett W Israelsen and Nisar R Ahmed. 2019. "Dave... I can assure you... that it's going to be all right..." A definition, case for, and survey of algorithmic assurances in human-autonomy trust relationships. *ACM Computing Surveys (CSUR)* 51, 6 (2019), 1–37.

[15] David B Kaber and Mica R Endsley. 2004. The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science* 5, 2 (2004), 113–153.

[16] Ramsundar Kalpagam Ganesan, Yash K. Rathore, Heather M. Ross, and Heni Ben Amor. 2018. Better Teaming Through Visual Cues: How Projecting Imagery in a Workspace Can Improve Human-Robot Collaboration. *IEEE Robotics & Automation Magazine* 25, 2 (June 2018), 59–71. https://doi.org/10.1109/MRA.2018.2815655

[17] Wansoo Kim, Luka Peternel, Marta Lorenzini, Jan Babič, and Arash Ajoudani. 2021. A Human-Robot Collaboration Framework for Improving Ergonomics During Dexterous Operation of Power Tools. *Robotics and Computer-Integrated Manufacturing* 68 (2021), 102084.

[18] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* 267 (2019), 1–38.

[19] Andrew H Mishkin, Daniel Limonadi, Sharon L Laubach, and Deborah S Bass. 2006. Working the Martian night shift-the MER surface operations process. *IEEE robotics & automation magazine* 13, 2 (2006), 46–53.

[20] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*. 220–229.

[21] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. 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 (2000), 286–297.

[22] David Puljiz, Erik Stöhr, Katharina S. Riesterer, Björn Hein, and Torsten Kröger. 2019. General Hand Guidance Framework using Microsoft HoloLens. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 5185–5190. https://doi.org/10.1109/IROS40897.2019.8967649 ISSN: 2153-0866.

[23] Christopher Reardon, Kevin Lee, John G. Rogers, and Jonathan Fink. 2019. Communicating via Augmented Reality for Human-Robot Teaming in Field Environments. In *2019 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. 94–101. ISSN: 2475-8426.

[24] Thomas B Sheridan. 2016. *Modeling Human–System Interaction: Philosophical and Methodological Considerations, with Examples*. John Wiley & Sons.

[25] Thomas B Sheridan. 2018. Comments on "Issues in human–automation interaction modeling: presumptive aspects of frameworks of types and levels of automation" by David B. Kaber. *Journal of Cognitive Engineering and Decision Making* 12, 1 (2018), 25–28.

[26] Aaquib Tabrez, Shivendra Agrawal, and Bradley Hayes. 2019. Explanation-based reward coaching to improve human performance via reinforcement learning. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 249–257.

[27] Aaquib Tabrez, Matthew B Luebbers, and Bradley Hayes. 2020. Automated failure-mode clustering and labeling for informed car-to-driver handover in autonomous vehicles. *arXiv preprint arXiv:2005.04439* (2020).

[28] Aaquib Tabrez, Matthew B Luebbers, and Bradley Hayes. 2020. A Survey of Mental Modeling Techniques in Human–Robot Teaming. *Current Robotics Reports* (2020), 1–9.

[29] Michael E. Walker, Hooman Hedayati, and Daniel Szafir. 2019. Robot teleoperation with augmented reality virtual surrogates. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI '19)*. IEEE Press, Daegu, Republic of Korea, 202–210.

[30] Chris Wickens, Gaia Dempsey, Andrew Pringle, Lucas Kazansky, and Stefanie Hutka. 2018. Developing and Evaluating an Augmented Reality Interface to Assist the Joint Tactical Air Controller by Applying Human Performance Models. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 62. SAGE Publications Sage CA: Los Angeles, CA, 686–690.

[31] A. W. W. Yew, S. K. Ong, and A. Y. C. Nee. 2017. Immersive Augmented Reality Environment for the Teleoperation of Maintenance Robots. *Procedia CIRP* 61 (Jan. 2017), 305–310. https://doi.org/10.1016/j.procir.2016.11.183