

Fiscal Year:	FY 2024	Task Last Updated:	FY 01/07/2024
PI Name:	Selva, Daniel Ph.D.		
Project Title:	HCAAM VNSCOR: Virtual Assistant for Spacecraft Anomaly Treatment During Long Duration Exploration Missions		
Division Name:	Human Research		
Program/Discipline:			
Program/Discipline-- Element/Subdiscipline:			
Joint Agency Name:		TechPort:	Yes
Human Research Program Elements:	(1) HFBP :Human Factors & Behavioral Performance (IRP Rev H)		
Human Research Program Risks:	(1) HSIA :Risk of Adverse Outcomes Due to Inadequate Human Systems Integration Architecture		
Space Biology Element:	None		
Space Biology Cross-Element Discipline:	None		
Space Biology Special Category:	None		
PI Email:	dselva@tamu.edu	Fax:	FY
PI Organization Type:	UNIVERSITY	Phone:	607-255-6351
Organization Name:	Texas A&M University		
PI Address 1:	Aerospace Engineering Department		
PI Address 2:	701 Ross St 3141 TAMU		
PI Web Page:			
City:	College Station	State:	TX
Zip Code:	77843-0001	Congressional District:	17
Comments:			
Project Type:	Ground	Solicitation / Funding Source:	2017-2018 HERO 80JSC017N0001-BPBA Topics in Biological, Physiological, and Behavioral Adaptations to Spaceflight. Appendix C
Start Date:	03/06/2019	End Date:	08/31/2025
No. of Post Docs:	0	No. of PhD Degrees:	0
No. of PhD Candidates:	2	No. of Master' Degrees:	0
No. of Master's Candidates:	1	No. of Bachelor's Degrees:	0
No. of Bachelor's Candidates:	1	Monitoring Center:	NASA JSC
Contact Monitor:	Whitmire, Alexandra	Contact Phone:	
Contact Email:	alexandra.m.whitmire@nasa.gov		
Flight Program:			
Flight Assignment:	NOTE: End date changed to 08/31/2025 per A. Beitman/JSC (Ed., 9/19/24) NOTE: End date changed to 02/28/2025 per A. Beitman/JSC (Ed., 1/25/23)		
Key Personnel Changes/Previous PI:			
COI Name (Institution):	Diaz Artilles, Ana Ph.D. (Texas A&M Engineering Experiment Station) Dunbar, Bonnie Ph.D. (Texas A&M Engineering Experiment Station) Wong, Raymond Ka Wai Ph.D. (Texas A & M, College Station)		
Grant/Contract No.:	80NSSC19K0656		
Performance Goal No.:			
Performance Goal Text:			

Task Description:	<p>This task is part of the Human Capabilities Assessments for Autonomous Missions (HCAAM) Virtual NASA Specialized Center of Research (VNSCOR).</p> <p>The research objective of this proposal is to investigate the impact of using Virtual Assistants (VA) to support crewmembers in the context of anomaly treatment during Long Duration Exploration Missions (LDEM), when ground support will be limited. A VA will be developed building upon the software architecture from existing VAs developed by the Principal Investigator (PI) for similar purposes. The VA will provide support for various aspects of anomaly treatment, including detecting and diagnosing the anomaly, as well as recommending a course of action. It will also have the ability to take initiative in the dialog with the user (mixed-initiative mode), and the ability to provide explanations for its actions. The impact of the VA on performance, cognitive workload, situational awareness, and trust, will be assessed through a set of three experiments with human subjects in a laboratory environment. The first experiment will establish the baseline impact (master-slave, no explanations), and subsequent experiments will study the effect of switching to the mixed-initiative mode and adding explanations. The system will also be deployed and tested in the Human Exploration Research Analog (HERA) analog environment.</p>
Rationale for HRP Directed Research:	<p>This project will provide standards and guidelines that will help NASA design similar virtual assistants to support astronauts during future long duration exploration missions. Such standards and guidelines will concern both the functionality and the user interface of the virtual assistant.</p>
Research Impact/Earth Benefits:	<p>In Year 5 of this project, we have worked primarily on Specific Aims 2 (Enhanced VA with self-explaining abilities) and 3 (validation in analog), although we have also worked on revisions for the journal paper corresponding to Specific Aim 1 (validation of the baseline agent in a lab environment).</p> <p>Concerning Specific Aim 2, we concluded Lab Experiment 2 on the effect of explanations and have submitted the corresponding two journal papers. This experiment addresses the following research question: How do explanations affect human performance, trust, cognitive workload (CW), situational awareness (SA), user satisfaction, and self-confidence for different levels of agent accuracy and uncertainty in human-AI collaborative anomaly diagnosis? The first paper focuses on the effects of explanations for different levels of agent accuracy, while the second paper focuses on the effects of uncertainty.</p> <p>The protocol for this experiment was as follows. Subjects start with a background survey and test and then they proceed to do two sessions, one in each explanation condition (order counterbalanced). In each session they work on 8 anomaly scenarios, 4 with high uncertainty and 4 with low uncertainty, in random order. After each anomaly, subjects fill out a confidence survey, a Jian survey, and a satisfaction survey. After each session, they also fill out a NASA Task Load Index (TLX), a Situational Awareness Rating Technique (SART) questionnaire, another Jian survey, and another satisfaction survey.</p> <p>30 subjects were recruited and performed the experiment. The major findings from the experiment are as follows:</p> <p>Effect of explanations (within-subjects comparison using surveys done after each condition) 1. Explanations improve trust ($p<0.001$) 2. Explanations improve #anomalies correctly diagnosed ($p=0.0039$), but do not significantly change the time to diagnosis ($p=0.37$) 3. Explanations improve SA ($p=0.009$) 4. Explanations slightly increase CW, but effect is not significant ($p=0.108$) 5. Explanations improve user satisfaction ($p<0.001$) 6. Explanations improve user confidence ($p<0.001$)</p> <p>Effect of uncertainty (within-subjects comparison using surveys done after each anomaly) 1. (High) Uncertainty decreases trust ($p<0.001$) 2. (High) Uncertainty decreases #anomalies correctly diagnosed ($p<0.001$) and time to diagnosis ($p<0.001$) 3. (High) Uncertainty decreases user satisfaction ($p<0.001$) 4. (High) Uncertainty decreases user confidence ($p<0.001$)</p> <p>The effect of uncertainty on CW and SA could not be measured due to the experimental design – they were only measured once with explanations and once without, averaging over the effects of uncertainty in both cases.</p> <p>The effects of accuracy were not significant, presumably due to individual differences leading to large variability between subjects. The only exception is that accuracy significantly improved the number of anomalies correctly diagnosed ($p<0.001$). The interaction effect between explanations and accuracy was also not significant.</p> <p>Concerning Specific Aim 3, we completed the C6 campaign and prepared for the C7 campaign.</p> <p>Results from the HERA C6 campaign: The major findings from the HERA C6 campaign ($N=16$) are as follows: • All subjects correctly resolved all scenarios. • Primary results did now show significant effects of the VA on any metrics. • The only exception is that attentional demand (a component of SA) was significantly higher with VA than without ($n=16$, $V=21$, $p=0.03$). • For group scenarios only, CW was higher with Daphne ($t(15) = -6.0207$, $p < 0.001$).</p> <p>Some of the insights we got from the exit interviews are as follows. The crew generally exhibited strong interest in using VAs for anomaly resolution, and enjoyed using Daphne. Many subjects showed an interest in the social aspects of VA and mentioned that they “attributed a personality to Daphne” and “talked about her as if she were another crewmember”. They all mentioned establishing trust very quickly once and for all thanks to Daphne “getting it right the first 3 times or so”. Almost none of them found the question answering capabilities essential because they were “going for speed” and didn’t feel like they needed to ask any questions. However, many subjects mentioned that question answering would be very useful in cases where Daphne recommended more than one diagnosis with the same confidence level. Moreover, crewmembers expressed an interest in the more interactive diagnosis and advanced explanations capabilities we are currently developing as something that would “significantly increase the usefulness” of the tool. Finally, they all confirmed that the scenarios generally felt very easy to diagnose and adding some more complexity would make it more interesting and fun. Note that detailed data regarding the interactions between the crewmembers and Daphne has not been analyzed yet.</p> <p>Preparation of HERA C7: In C7, we will test a more advanced version of the Daphne agent with improved self-explaining abilities while also testing more complex anomaly scenarios in which there are cascading and simultaneous failures. The timeline for the campaign is shown below. • Phase 1 – Group sessions to study the effect of time delay (crew allowed to talk to MCC) – MD1-4: Group study with and without VA at 0s time delay – 2x1h –</p>
Task Progress:	<p>Results from the HERA C6 campaign: The major findings from the HERA C6 campaign ($N=16$) are as follows: • All subjects correctly resolved all scenarios. • Primary results did now show significant effects of the VA on any metrics. • The only exception is that attentional demand (a component of SA) was significantly higher with VA than without ($n=16$, $V=21$, $p=0.03$). • For group scenarios only, CW was higher with Daphne ($t(15) = -6.0207$, $p < 0.001$).</p> <p>Some of the insights we got from the exit interviews are as follows. The crew generally exhibited strong interest in using VAs for anomaly resolution, and enjoyed using Daphne. Many subjects showed an interest in the social aspects of VA and mentioned that they “attributed a personality to Daphne” and “talked about her as if she were another crewmember”. They all mentioned establishing trust very quickly once and for all thanks to Daphne “getting it right the first 3 times or so”. Almost none of them found the question answering capabilities essential because they were “going for speed” and didn’t feel like they needed to ask any questions. However, many subjects mentioned that question answering would be very useful in cases where Daphne recommended more than one diagnosis with the same confidence level. Moreover, crewmembers expressed an interest in the more interactive diagnosis and advanced explanations capabilities we are currently developing as something that would “significantly increase the usefulness” of the tool. Finally, they all confirmed that the scenarios generally felt very easy to diagnose and adding some more complexity would make it more interesting and fun. Note that detailed data regarding the interactions between the crewmembers and Daphne has not been analyzed yet.</p> <p>Preparation of HERA C7: In C7, we will test a more advanced version of the Daphne agent with improved self-explaining abilities while also testing more complex anomaly scenarios in which there are cascading and simultaneous failures. The timeline for the campaign is shown below. • Phase 1 – Group sessions to study the effect of time delay (crew allowed to talk to MCC) – MD1-4: Group study with and without VA at 0s time delay – 2x1h –</p>

	<p>MD5-10: Group study with and without VA at 5s time delay – 2x1h – MD11-12: Group study with and without VA at 1min time delay – 2x1h • Phase 2 – Individual sessions for main study (not allowed to talk to MCC) – Week 3: Individual training, 1h per crewmember, 2 simple scenarios, 1 with VA and 1 without – Weeks 4-6: 2 sessions per crewmember per week, 1 with VA and 1 without, 3x4x2x1h=24h</p> <p>As shown above, the group scenarios from C6 have been replaced with a shorter study on the effect of cislunar time delays (5sec) compared to no delays or longer delays (1 min). For these scenarios, the crew will also be allowed to communicate with MCC, to gain insight into the roles a VA can play for cislunar operators given that MCC is available albeit with a non-negligible time delay of a few seconds.</p> <p>The second phase is the main phase of individual scenarios and follows the same design as in C6, with the caveat that a (re)-training session is done for each crewmember in week 3 before they start their individual tests.</p> <p>Finally, in the next year, we also plan on finalizing the 3 publications under review (which should finalize work related to Specific Aims 1 and 2) and submitting a fourth one on the results of C6.</p>
Bibliography Type:	Description: (Last Updated: 02/21/2025)
Abstracts for Journals and Proceedings	<p>Selva D, Dutta P, Josan PK, Dunbar BJ, Wong KRW, Diaz-Artiles A. "Virtual assistant for anomaly resolution in long duration exploration missions: Overview of results so far and next steps." 2024 Human Research Program Investigators' Workshop, Galveston, Texas, February 13-16, 2024.</p> <p>Abstracts. 2024 Human Research Program Investigators' Workshop, Galveston, Texas, February 13-16, 2024. , Feb-2024</p>
Abstracts for Journals and Proceedings	<p>Dutta P, Josan PK, Dunbar BJ, Wong KRW, Diaz-Artiles A, Selva D. "Effect of agent accuracy, confidence, and transparency on the evolution of trust in AI-assisted anomaly diagnosis." 2024 Human Research Program Investigators' Workshop, Galveston, Texas, February 13-16, 2024.</p> <p>Abstracts. 2024 Human Research Program Investigators' Workshop, Galveston, Texas, February 13-16, 2024. , Feb-2024</p>
Abstracts for Journals and Proceedings	<p>Josan PK, Dutta P, Dunbar BJ, Wong KRW, Selva D, Diaz-Artiles A. "Results from Hera analog testing of a virtual assistant for anomaly resolution." 2024 Human Research Program Investigators' Workshop, Galveston, Texas, February 13-16, 2024.</p> <p>Abstracts. 2024 Human Research Program Investigators' Workshop, Galveston, Texas, February 13-16, 2024. , Feb-2024</p>