Investigating Human-AI Team Fluency in Autonomous Medical Evacuation: A Study of Novice Aviator Cognitive States and Human-AI Interface Design

Sanya Doda*
Georgia Institute of Technology, Aerospace Engineering Department. Atlanta, GA.

Richard Agbeyibor†
Georgia Institute of Technology, Aerospace Engineering Department. Atlanta, GA.

Carmen Jimenez Cortes‡
Georgia Institute of Technology, Electrical Engineering Department. Atlanta, GA.

Jack Kolb§
Georgia Institute of Technology, Electrical Engineering Department. Atlanta, GA.

José M. Magalhães Júnior ¶
Georgia Institute of Technology, Aerospace Engineering Department. Atlanta, GA.

Karen M. Feigh ‖
Georgia Institute of Technology, Aerospace Engineering Department. Atlanta, GA.

I. Nomenclature

AP	=	Autonomous Pilot
$DARPA$	=	Defense Advanced Research Projects Agency
EMT	=	Emergency Medical Technician
HAI	=	Human-AI Interface
$HEMS$	=	Helicopter Emergency Medical Operations
ISR	=	Intelligence, Surveillance, and Reconnaissance
$MEDEVAC$	=	Medical Evacuation
$MSFS$	=	Microsoft Flight Simulator
ONR	=	Office of Naval Research

II. Introduction

Autonomous technologies are primed to revolutionize the aerospace industry domain by influencing the way aerial vehicles operate. Additionally, the relative scarcity of human pilots and the administrative challenges of securing a prompt service in an emergency further support the development for autonomous aircraft for certain missions. While it appears that fully automated helicopters will ultimately be used for a range of missions in the next few decades, often many missions require both a pilot and specialist crew members in the back, e.g. medical evacuations (MEDEVAC). Developing autonomously piloted rotorcraft may still require require human involvement through the provision of essential healthcare caregivers. This shift to autonomous systems makes it imperative to evaluate how effectively the AI pilot and such specialized human rear-crew members can collaborate.

*PhD Student, sdoda3@gatech.edu.
†PhD Student, richard.agbeyibor@gatech.edu.
‡PhD candidate, ccortes9@gatech.edu.
§PhD candidate, kolb@gatech.edu.
¶Research Engineer II, AIAA Senior Member. E-mail: jose.magalhaes@gatech.edu.
‖Professor and Associate Chair for Research, AIAA Associate Fellow. E-mail: karen.feigh@gatech.edu.

This research analyzes a future medical transportation scenario in which medical personnel with little to no aviation background operate aboard autonomously flown helicopters, emphasizing on two key goals:

- Finding the critical elements influencing mission effectiveness and human-AI interaction under off-nominal circumstances.
- Evaluating the efficiency of the interface in facilitating human-AI interaction.

The cognitive and interacting aspects of human-machine teaming in HEMS operations are examined in this study using a high-fidelity simulation approach which incorporated eye-tracking technologies and realistic medical scenarios. We analyze this problem using three crucial perspectives that are in line with the recent developments in human-machine collaboration:

Human Factors: Examining medical professionals' cognitive load, situational awareness, and decision-making styles in situations involving multiple tasks and off-nominal circumstances

Interface Design: Assessing how well a specifically developed human-AI interface facilitates interaction between the two agents

Team Fluency: Examining how human cognitive processes, their ability to perform tasks and overall mission success are interdependent in off-nominal circumstances

For autonomous aerospace applications where human expertise and autonomous systems capabilities must be efficiently merged, the findings of this investigation could carry significant implications for the design of human-machine interfaces, training guidelines, and operating practices.

III. Background & Literature Review

Autonomous flight has been an end goal of many commercial and military transport organizations for decades. Indeed, the history of aviation has shown that reductions in personnel are possible, but with significant engineering and development effort. On early transport airplanes, a flight crew of five members was the norm: two pilots, a flight engineer, a navigator, and a radio operator. Advanced avionics and automation technologies have significantly transformed cockpit operations, resulting in a gradual reduction of the crew members on-board [1]. In present days, transport airplanes equipped with high technology systems can be operated with a flight crew with two members: captain and co-pilot, known as Dual-Pilot Operations (DPO) comprising a Pilot Flying (PF) and a Pilot Monitoring (PM). Since the aeronautical industry is continuously looking for more efficient aircraft, researchers have investigated and discussed the feasibility of implementing Single-Pilot Operations (SPO) in traditional two-pilot contexts, such as FAR (Federal Aviation Regulations) Part 121 operations due to its potential for cost savings [2–5].

With the advance of autonomous vehicles and AI-based intelligent systems, researchers have investigated the operation of fully autonomous aircraft and reduced pilot operations and to understand how the new technology may affect the behavior of the passengers. For instance, in [6], the authors discuss the use of Pilotless Aircraft (PA) and present the results of an investigation into attitudes towards and willingness to fly in PA among a sample of 711 UK people known to fly at least occasionally. The authors concluded that risk, excitement, and innovation are the three primary components that significantly influence willingness to fly in a PA.

Finally, the recent AACUS project by the Office of Naval Research (ONR) has demonstrated that autonomous flight of rotorcraft is possible and even likely in the few decades under specialized circumstances. AACUS – the Autonomous Aerial Cargo Utility System – a modular system consisting of sensors and software, represents a significant advancement in autonomous flight capability for rotary-wing aircraft. Utilizing this versatile package kit, rotorcraft platforms of distinct categories could be equipped to carry out complicated tasks autonomously, such as flight planning, obstacle avoidance, and unplanned landing site selection, even within adverse terrain [7]. Further, the Defense Advanced Research Projects Agency (DARPA) has partnered with Sikorsky to develop an autonomy kit that can be placed on any rotary-wing platform and provide it with an autonomous capability. Using this autonomy kit, Sikorsky and DARPA demonstrated in 2024 how an optionally piloted Black Hawk helicopter can be flown and controlled by an operator in the cabin or on the ground just by entering high level mission goals [8]. In previous work [9], we presented the results of an empirical human-AI collaboration study between non-pilot human crewmates and an Autonomous Pilot to accomplish an Intelligence, Surveillance, and Reconnaissance (ISR) mission. The results indicate that it is possible to reduce crew requirements on missions like ISR which currently require both a crew of pilots and a crew of mission specialists.

IV. Methodology

A. Experimental Apparatus

To accurately assess participant performance and situation awareness in simulated air medical transport scenarios, create a realistic simulation environment was created. This included the careful construction of a high-fidelity cabin simulator that closely resembled a rotorcraft's aft section. To enhance the ecological relevance, various crucial components necessary for a medical evacuation operation were included in the experimental setting. Multiple screens showing significant mission-critical data, such as patient vital signs, communication and navigation interface, backup helicopter gauges, along with a window-view simulation, were made available to participants. Additionally, to replicate the specific tools accessible to medical personnel in real-world in-flight situations, the cabin was equipped with components such as a surgical kit, vital medical equipment including IV systems, and a life-like baby doll representing a pediatric patient.

Fig. 1 Cabin resembling the rear compartment of the aircraft

B. Simulation and Control Technologies

The study made use of the Microsoft Flight Simulator (MSFS) for high-resolution, out-the-window visuals and auditory replication to boost realistic portrayal and immersive experience. As creating and deploying an actual fully automated pilot was beyond the scope of the study, a Matlab-Simulink based simulator was employed to simulate the fundamental helicopter dynamics and control techniques. The use of a derivative nominal controller allows the vehicle to reach the desired destination, at a nominal speed. The helicopter behaved nominally unless explicitly instructed by the experimental protocol to malfunction. Safety and operational restrictions were ensured by integrating Control Barrier Functions (CBF) [10]. This method stopped the rotorcraft from flying over the destination helipad Without explicit human consent. The CBF applied a dynamic slowdown and possible stop mechanism until the medic completed a crucial landing checklist that confirmed the availability of helipad, protection of the patient and participants' personal safety

C. Human-AI (HAI) Interface Design

A Human-AI (HAI) interface was specifically designed to enable smooth interaction between the AI pilot and the medic, offering interactive features and contextual awareness. One essential element was the map, which provided participants with real-time information into the position of the rotorcraft, helipads in the vicinity that were clearly marked to distinguish between commercial and the ones with medical facilities and the estimated time of arrival.

In addition to providing navigational information, the interface enabled participants to interact directly with the simulated autonomous pilot (AP). Medical professionals could make a mid-flight destination change request or ask the

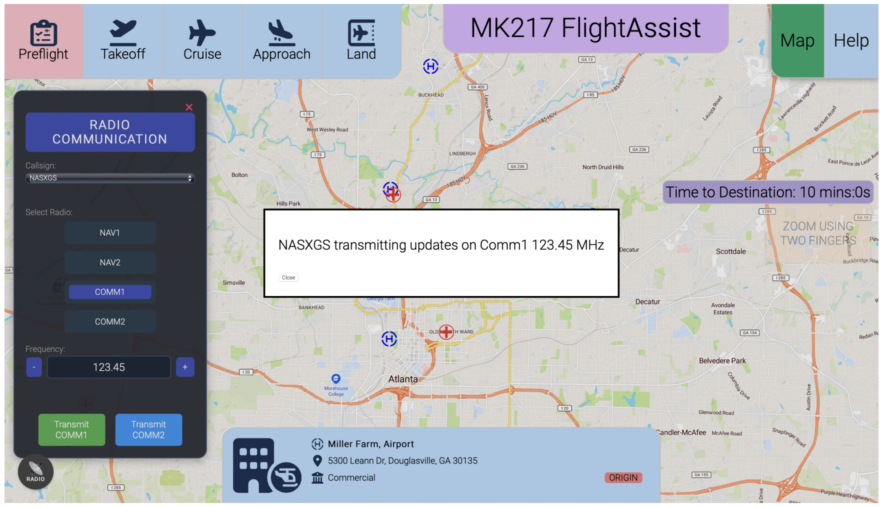

Fig. 2 HAI Interface showing the map and radio feature

AP to initiate emergency landing protocols.The proactive approach to safety management was among the crucial aspects of the interface. Prior to departure, the AP would ask the medic to go through a safety checklist aimed at affirming the use of seatbelt and the safety of the patient.

Jarvis, a voice-assistant, was introduced to improve the interaction between the medic and the AP especially in off-nominal and high-workload situations. The intent for the voice assistant was to, call the medic's immediate attention when circumstances get critical and facilitate interaction even when the medic's hands and visual attention were engaged with medical duties.This means of communication enabled medical professionals to initiate voice instructions (such as "Jarvis, change destination") and get multi-modal assistance (voice cues and visual displays) while continuing to have the final decision-making authority

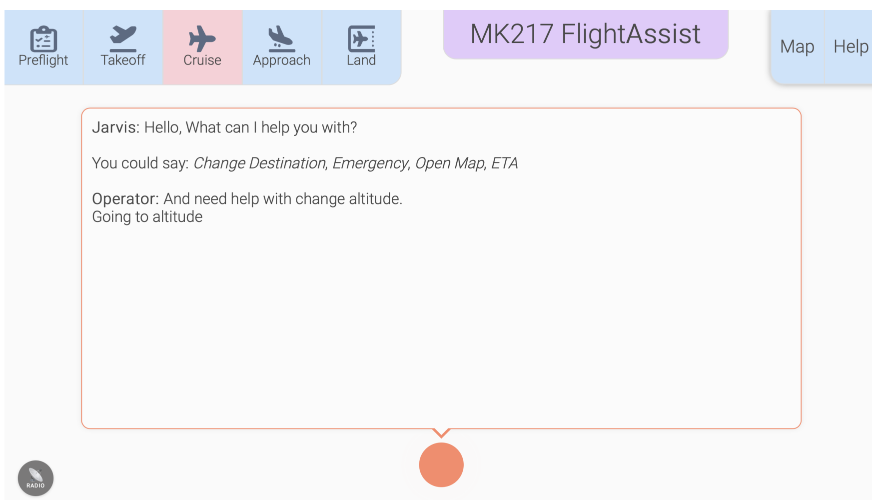

Fig. 3 Voice Assistant *Jarvis* in active state

D. Scenario Design

Four different scenarios were crafted in order to accurately assess participant responses in a range of real-life circumstances:

Training Scenario: Intended to acquaint users with response protocols, task navigation, and system interactions.

Medical Emergency Scenario: Participants were required to administer medication and interact with the AI pilot to adjust the altitude.

Fuel Gauge Scenario: Participants were presented with a possibly erroneous fuel gauge reading and were asked to decide based on their judgment, AI suggestion and the accessible ground controller consultation.

Complex Multi-Event Scenario: This scenario tests key decision-making under a tight time constraint by simulating multiple off-nominal incidences involving sensor malfunction, failure of the engine, and decline in patients health.

With the exception of the training scenario, situations were given in a randomized order to mitigate the likelihood of order effects. The training scenario was designed to be the least taxing and the Complex Multi-Event Scenario to be the most taxing.

E. Participant Tasks

The participants were allocated with primary and secondary tasks that mirrored actual medical evacuation duties. Their main task was to make sure the patient remains stable during flight and is transported safely. The Secondary assignments included recording the patient's vital signs periodically using a vitals logger interface and give the ground controller regular updates regarding the flight status, the patient status and the approximate time of arrival via radio.

F. Data Collection Methods

The study collected data using an elaborate mixed-methods approach. This included conducting Pre-trial, in between scenarios and post-trial surveys along with semi-structured interview to gauge participant perceptions, trust, situational awareness, and cognitive demands. Quantitative data on error rates, human-AI interactions, and secondary task completion times were recorded Furthermore, qualitative insights were gathered from interface interactions logs, participants behavior and observational data. Using eye-tracking technologies, fixation duration, gaze transitions and attention allocation between interfaces data was recorded to study the distribution of visual attention and obtain comprehensive details on patterns of interaction and cognitive processes.

G. User Study

In order to examine human-AI collaboration in medical evacuation situations, the study enlisted a varied group of 22 medical specialists. The pool comprised of 11 Certified Emergency Medical Technicians (EMTs), 6 Certified Nurses 2 Certified Medical Professionals, 1 Certified Paramedic and remaining 4 being healthcare providers with hospital experience. The participant age ranged between 18 and 45 years. The recruited participants had minimal experience with piloting aircraft and autonomous systems. Only 5 participants claimed they had moderate to high levels of AI experience. Nearly all participants (20 of 22) had experience with medical evacuation tasks.

Throughout the entirety of the study planning and execution process, ethical considerations were given utmost importance. Before recruiting, permission for the study from Institutional Review Board (IRB) was obtained. Recruited participants were given comprehensive details on the goals, potential risks, and benefits of the study as part of an extensive informed consent procedure. They were made fully aware of their right to participate voluntarily and their freedom to leave at any time without facing any repercussions. By using strict anonymization procedures, such as assigning unique identifying codes and removing all personally identifiable information, data management practices placed a high priority on participant confidentiality. Secure data processing and storage were ensured by the implementation of strict confidentiality agreements. Only approved members of the study team could access the encrypted data, and aggregated reporting techniques were used.

V. Preliminary Results and Discussion

The preliminary investigation of the MEDEVAC simulation yielded substantial findings into human-AI collaboration and task performance under differing scenario complexity. The effectiveness of our experimental design to generate varied workload across situations has been demonstrated by NASA TLX investigation. Quantitative investigation utilizing the Analysis of Variance (ANOVA) indicated that off-nominal occurrences had a statistically significant impact on secondary task completion times [$F(3,71) = 6.26$, $p < 0.001$]. Complex performance differences between scenarios were revealed by post-hoc Tukey's analysis. The Medical Emergency scenario showed considerably faster vital signs logging times than the Training scenario ($p = 0.009$), while the Multiple Emergencies scenario showed significantly slower vital signs logging times than both the Aviation Emergency ($p = 0.044$) and Medical Emergency scenarios ($p = 0.002$).

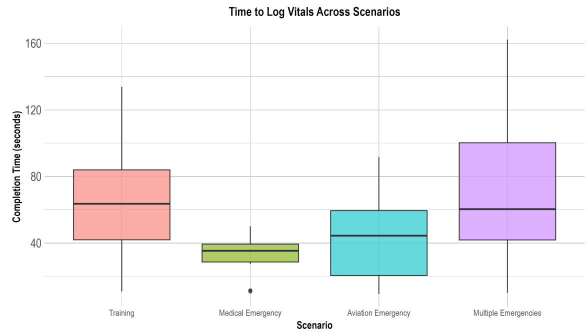

Fig. 4 Time taken to log vitals

The number of missed secondary tasks was found to be consistent with increased scenario complexity, indicating that participants were still able to complete mission-critical activities despite increased cognitive demands.

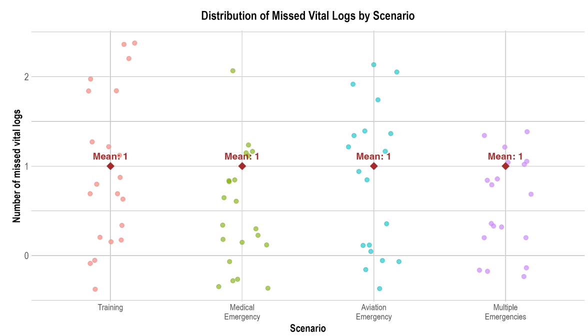

Fig. 5 Number of missed vital logging

The Coefficient of Variance (CV), which measures inter-participant performance variability, provided important details on performance consistency. There was notable performance variation in complicated, multi-emergency circumstances, with the Medical Emergency scenario showing the least amount of performance variability (CV = 34.7%) and the Multiple Emergencies scenario showing the most variability (CV = 61.6%).

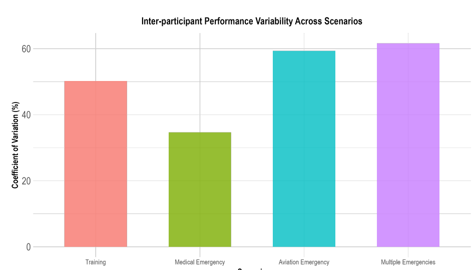

Fig. 6 Inter-participant Variability

A noteworthy pattern of behavior in relation to task prioritizing was observed. Medical activities were regularly given precedence over aviation-related responsibilities, even when the latter would have jeopardized flight safety. Longer emergency reaction times in Medical Emergency scenario and lower Coefficient of Variation metrics support this finding.

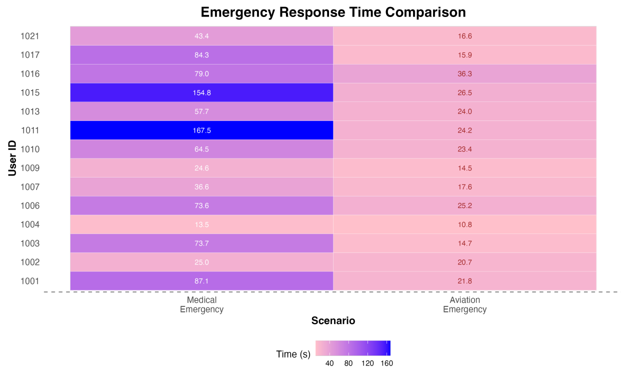

Fig. 7 Emergency Response Time

75% of participants used the voice help interface (Jarvis), typically when they were unsure how to carry out an action but rarely when they were felt overworked. Participants with less technology knowledge preferred manual task completion to voice-assisted interactions, indicating that technological understanding had an impact on interface interaction. Post-trial interviews revealed that although users occasionally missed to notice the voice feature's presence in multi-task circumstances, it appeared useful in offering task reminders and guidance during crises. Furthermore, the success of the experimental design in creating cognitively demanding scenarios was confirmed by participants who reported experiencing higher stress levels when faced with multiple emergencies simultaneously

VI. Conclusion

Preliminary data analysis reveals that, while medical personnel may work with an AP to handle crises, they may not be best positioned to prioritize aircraft-related activities during critical incidents. These results highlight how crucial human elements are when creating autonomous rotorcraft systems, especially for missions that need to manage both aviation and medical tasks at the same time. Various additional aspects of study, including an exhaustive analysis of the survey data to get a better understanding of how participant views, trust levels, and situational awareness evolved throughout the experiment, accuracy and time taken for radio updates to clarify if off-nominal occurrences have an effect on communication styles and information transfer and eye-tracker data analysis to study the distribution of visual attention are currently under investigation. The results presented underline the need for more study on human-AI interaction and point to possible directions for focused training, enhanced system architecture, and improved interface design. Understanding the complex dynamics of human-AI collaboration is becoming more and more important as autonomous technologies develop in order to guarantee safe and effective operating procedures. As we transition to more autonomous technology, human factors considerations are crucial for safe and efficient operations. We anticipate developing a more comprehensive understanding of human-AI interaction as we carry out more analysis.

Acknowledgments

This work was funded in part by the Office of Naval Research, Science of Autonomy grant N00014-21-1-2759 Human-AI Collaboration in Autonomous Aerial Vehicles. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the U.S. Navy, Department of Defense, or the U.S. Government.

References

[1] Puca, N., and Guglieri, G., "Enabling Civil Single-Pilot Operations: A State-of-the-Art Review," *Aerotecnica Missili & Spazio*, 2024. URL `https://api.semanticscholar.org/CorpusID:270498673`.

[2] Comerford, D. P., Brandt, S. L., Lachter, J., Wu, S.-C., Mogford, R., Battiste, V., and Johnson, W. W., "NASA's Single-Pilot Operations Technical Interchange Meeting: Proceedings and Findings," 2013. URL `https://api.semanticscholar.org/CorpusID:108861053`.

[3] Lim, Y., Bassien-Capsa, V., Ramasamy, S., Liu, J., and Sabatini, R., "Commercial airline single-pilot operations: System

design and pathways to certification," *IEEE Aerospace and Electronic Systems Magazine*, Vol. 32, No. 7, 2017, pp. 4–21. https://doi.org/10.1109/MAES.2017.160175.

[4] Vu, K.-P. L., Lachter, J., Battiste, V., and Strybel, T. Z., "Single Pilot Operations in Domestic Commercial Aviation," *Human Factors*, Vol. 60, No. 6, 2018, pp. 755–762. https://doi.org/10.1177/0018720818791372, URL https://doi.org/10.1177/0018720818791372, pMID: 30063410.

[5] Piera, M. A., Muñoz, J. L., Gil, D., Martin, G., and Manzano, J., "A Socio-Technical Simulation Model for the Design of the Future Single Pilot Cockpit: An Opportunity to Improve Pilot Performance," *IEEE Access*, Vol. 10, 2022, pp. 22330–22343. https://doi.org/10.1109/ACCESS.2022.3153490.

[6] Bennett, R., and Vijaygopal, R., "Air passenger attitudes towards pilotless aircraft," *Research in Transportation Business and Management*, Vol. 41, 2021, p. 100656. https://doi.org/https://doi.org/10.1016/j.rtbm.2021.100656, URL https://www.sciencedirect.com/science/article/pii/S2210539521000390.

[7] Jones, S., "Autonomous Helicopter Technology Wins Another Major Award - Composites Opportunities in eVTOL," , Mar. 2022. URL www.nre.navy.mil/media-center/news-releases/autonomous-helicopter-technology-wins-another-major-award.

[8] HARPER, J., "Autonomous Helicopters Seen as Wave of the Future," *National Defense*, Vol. 102, No. 771, 2018, pp. 26–27.

[9] Agbeyibor, R., Ruia, V., Kolb, J., and Feigh, K. M., "Joint Intelligence, Surveillance, and Reconnaissance Mission Collaboration with Autonomous Pilots," *Human Factors and Ergonomics Society Annual Meeting*, 2024. https://doi.org/10.1177/10711813241262302, URL https://journals.sagepub.com/doi/10.1177/10711813241262302.

[10] Ames, A. D., Coogan, S., Egerstedt, M., Notomista, G., Sreenath, K., and Tabuada, P., "Control barrier functions: Theory and applications," *2019 18th European control conference (ECC)*, IEEE, 2019, pp. 3420–3431.