Human-AI Interaction in Autonomous Medical Evacuation Helicopters

Sanya Doda	Richard Agbeyibor	Carmen Jimenez Cortes
sdoda3@gatech.edu	richard.agbeyibor@gatech.edu	ccortes9@gatech.edu
PhD Student	PhD Student	PhD Candidate
Georgia Institute of Technology	Georgia Institute of Technology	Georgia Institute of Technology
Jack Kolb	**Jose M. Magalhaes Jr**	**Karen M. Feigh**
kolb@gatech.edu	jmmj6@gatech.edu	karen.feigh@gatech.edu
PhD Candidate	Research Engineer	Professor
Georgia Institute of Technology	Georgia Institute of Technology	Georgia Institute of Technology

INTRODUCTION

The vertical flight industry is on its way to a transformative era, with autonomous technologies set to alter aerial vehicle operations. While it seems certain that fully autonomous helicopters will eventually be deployed for a variety of missions, some high-stakes situations—like medical evacuations (MEDEVAC)—will for the foreseeable future demand human participation in the form of Emergency Medical Care-giving Crew. A shortage of human pilots, combined with the logistical constraints of ensuring a rapid response in emergencies, further pushes the case for autonomous rotorcraft.

This study investigates a hypothetical future situation in which a helicopter is autonomously piloted while a human medic manages patient care onboard. The primary research question is: Can medics without extensive training effectively collaborate with an autonomous pilot to ensure mission success, particularly during emergencies? Furthermore, it examines the feasibility of expecting these medics to respond to requests, to initiate requests, and to collaborate with the autononomous pilot during realistic missions under both nominal and off-nominal conditions.

By subjecting participants to a realistic MEDEVAC simulation where human medical personnel are placed in the cabin of an AI-piloted rotorcraft, we assessed their ability to respond to both aviation and medical emergencies and understand the nuances of this interaction, and its implications for future system design and operational protocols.

This abstract presents a comprehensive overview of our methodology, key findings, and their implications for the vertical flight community. By examining these critical human factors in the context of autonomous rotorcraft operations, we aim to examine whether investing in such hybrid systems can yield tangible benefits in terms of lives saved and operational efficiency and inform future design decisions, operational procedures, and regulatory frameworks.

BACKGROUND

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 (Ref. 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 (Refs. 2–5).

With the advance of autonomous vehicles and AI-based intelligent systems, researchers have investigated the operation of fully autonomous aircraft and how the new technology may affect the behavior of the passengers. For instance, in (Ref. 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 next decade or so. 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 (Ref. 7).

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 (Ref. 8).

In previous work (Ref. 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.

METHOD

In this experimental study, participants were immersed in a high-fidelity cabin simulator that replicated the rear compartment of a rotorcraft shown in Figure 1. The participant had access to multiple monitors displaying crucial information for the mission: the patient's vitals, the navigation interface, backup helicopter gauges, and a window. The experimental environment was augmented by placing a fictitious patient (infant) inside the cabin. Furthermore, the cabin was equipped with a medical kit and essential instruments such as IV systems, mirroring the resources available to medical professionals during in-flight emergencies. This framework strengthened the study's validity on ecological grounds and made it feasible to gauge participant performance and situational awareness in simulated air medical transport scenarios more precisely. For an enhanced immersive experience, we used Microsoft Flight Simulator (MSFS), which provided high quality graphics as well as sound. The helicopter dynamics and controller were implemented in a Matlab-Simulink based simulator. For the control strategy, we used a Derivative controller, commanding the helicopter towards the helipad at a nominal speed chosen by the participant. Additionally, we used Control Barrier Functions (CBFs) (Ref. 10) to prevent the helicopter from traversing over the destination helipad before the human on board approved it. The CBF around the destination helipad slowed the vehicle, and stopped it if necessary, until the medic on board completed the landing checklist. This checklist required the participant to verify that: **(1)** the helipad was available, **(2)** the patient was buckled in, and **(3)** the participant was buckled in. After successfully completing this checklist, the CBF was removed and the nominal controller drove the helicopter to the center of the helipad to conclude the mission. The CBF used around the helipad of coordinates (hp_x, hp_y) was:

$$h(x) = \|x - hp_x\|_2^2 - \|y - hp_y\|_2^2 - R^2, \tag{1}$$

where (x, y) are the Cartesian coordinates of the vehicle and R^2 is the radius around the helipad that the helicopter must stay outside of. This CBF is implemented in a minimally invasive manner, solving the quadratic program in (2), where α is a positive constant and $\hat{u}$ is the nominal derivative controller. This optimization problem is solved every 30 milliseconds, and the resulting u is applied to the system dynamics to obtain the updated latitude and longitude of the vehicle to be used in MSFS.

$$\begin{aligned} &\underset{u}{\text{minimize}} \quad ||u - \hat{u}||^2 \\ s.t. \quad &\dot{h} \geq -\alpha \cdot h \end{aligned} \tag{2}$$

In a Medical Evacuation scenario and under the assumption of having a crew not trained in aviation or autonomy, runtime assurance mechanisms such as CBFs allow us to guarantee a safe behavior of the autonomous system without relying on manual human control.

Participants were tasked with the primary objective of ensuring the safe transportation and stability of the infant patient throughout the simulated flight and two secondary tasks of regularly logging the patient's vital signs and providing status updates to the ground controller via radio. These tasks were designed to mirror the real-world responsibilities of medical personnel during air medical transport missions. To evaluate participant responses to various off-nominal events that could occur during real-world flights, four distinct scenarios were developed. The scenarios were presented in a randomized order (excluding the training scenario) to mitigate potential order effects. The scenarios included:

1. A training scenario to familiarize participants with system interactions, navigating tasks, and responding to different circumstances.
2. A medical emergency scenario requiring medication administration and communication with the AI pilot to request altitude adjustments.
3. A scenario involving a potentially false fuel gauge reading, requiring participants to make decisions based on AI recommendations and optional ground controller consultation.

Figure 1: Experimental Cabin - MEDEVAC Configuration

4. A complex scenario involving multiple simultaneous off-nominal events, including sensor malfunction, engine failure, and patient deterioration, necessitating critical decision-making under time pressure.

Data collection employed a mixed-methods approach, incorporating pre-trial, mid-trial, and post-trial surveys to assess participants' perceptions, trust levels, situational awareness, and workload. Additionally, observational and interface data were collected throughout the scenarios to capture participant behaviors, decisions, and interactions. Secondary task completion times, error rates, and the frequency of human-AI interactions were among the quantitative data collected. The distribution of visual attention was measured using eye tracking equipment. This comprehensive experimental design allowed for the examination of human-AI interaction in an operational context, providing insights into decision-making processes, communication patterns, and trust dynamics in autonomous medical evacuation scenarios.

Current Status

The data collection phase of the experiment is complete and data analysis has begun. A thorough examination of the collected data is currently being conducted by our study team. This section summarizes our work to date and preliminary results from the exploratory analysis we have carried out.

DATA COLLECTION AND ANALYSIS

The detailed log of all the interactions the medic engaged in while navigating through the simulated scenarios enabled the evaluation of the influence off-nominal events had on mission outcomes, participant behavior, and variations in human-AI interaction.

Task Performance Analysis

Using duration of completion, the accuracy of fulfillment, and the number of missed occurrences of the secondary tasks as the key indicators, we assessed how various off-nominal circumstances affected task performance.

Vitals Logging

1. **Time to Log Vitals:** For every scenario, a repeated measure ANOVA was performed on the amount of time needed to record vital signs. The results showed substantial differences across scenarios ($p < 0.05$). Significant distinctions in vitals logging time amongst scenarios 2 and 4 and scenarios 3 and 4 were revealed by post-hoc pairwise comparisons. Furthermore, a mixed model analysis indicated a considerable variation in task completion times between participants.

2. **Missed Vital Logs:** Performing exploratory data analysis, calculating mean and variance on the number of missed vital sign logs for each scenario, and applying the Poisson GLM model, demonstrate that the average number of missed vital signs logs was not affected by the nature of the scenario. This indicates the ability of participants to interact with the system and maintain the completion rate of the secondary task irrespective of the intricacy of off-nominal events.

ONGOING ANALYSIS

Various other elements of research such as the ones mentioned below are presently being investigated.

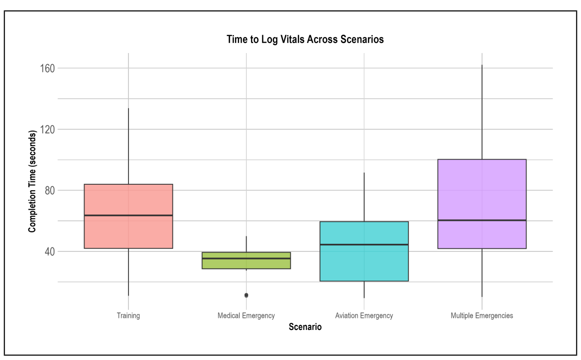
1. Completion of Radio Updates: An investigation of the duration and precision of radio updates continues to be underway. This serves to shed light on whether communication styles and the transfer of information are impacted by off-nominal events.
2. Patterns of Interaction: Extensive examination of user interactions with the autonomous system under various settings is currently in progress. This will assist in determining how the dynamics of human-AI collaboration are affected by off-nominal events.
3. Survey Data: Pre-, mid-, and post-trial survey data will also be analyzed as a means to better comprehend how participant perceptions, levels of trust, and situational awareness changed over the course of the experiment.

PRELIMINARY RESULTS AND DISCUSSION

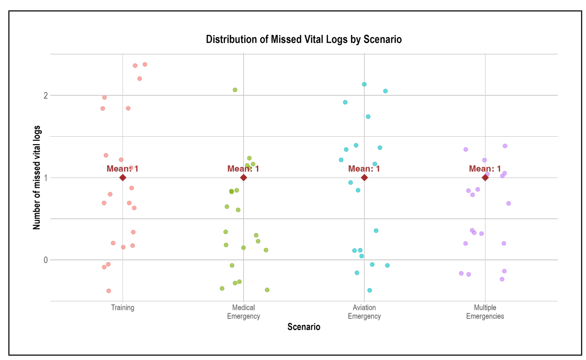
Our findings suggest a complex interplay between human decision-making, task prioritization, and interaction with the autonomous pilot. Notably, we observed that while the medics generally responded to aviation-related tasks as expected, their primary focus remained on their medical responsibilities. This raises important questions about the optimal division of responsibilities in such mixed-initiative missions. Furthermore, our study revealed intriguing variations in how individuals perceived and responded to different types of emergencies. Some participants demonstrated a heightened sense of urgency towards medical emergencies compared to aviation-related issues, highlighting the need for careful consideration in crew resource management and training protocols for these novel operational paradigms.

Preliminary results from exploratory data analysis indicate that:

- The secondary task's completion time tends to be substantially impacted by off-nominal events as suggested by the results of the ANOVA test conducted ($F(3,71) = 6.26$, $p < 0.001$). Figure 2a depicting the time taken by the participants to log the vitals in each scenario also highlights this result. Post-hoc Tukey's analysis demonstrated that, in comparison to the Training scenario, the Medical Emergency scenario had substantially faster vitals logging times ($p = 0.009$). Additionally, the time to log the vitals during the Multiple Emergencies scenario was notably slower than those of the Aviation Emergency ($p = 0.044$) and Medical Emergency ($p = 0.002$).
- As scenarios become more complicated, while the task completion times begin to increase and the task accuracy appears to diminish, the occurrences of missed secondary tasks do not seem to be affected (Figure 2b), suggesting that the complexity of the scenarios did not hinder the task completion and the overall success of the mission.



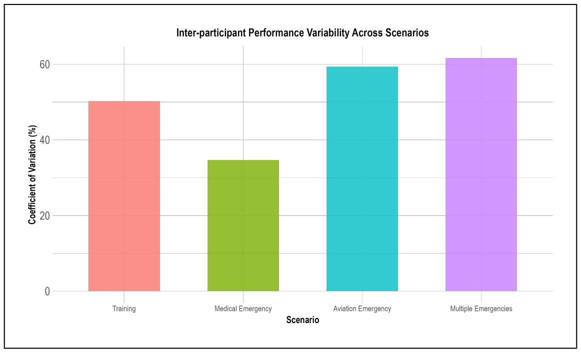
(a) Vitals Logging Time



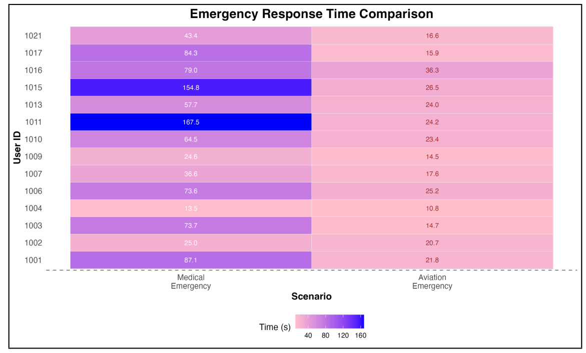
(b) Missed Vitals Logging

Figure 2: Vitals Logging

- Participant performance varies noticeably from one another, suggesting possible areas for focused training or system design enhancements. The Coefficient of Variation (CV) for completion time, used to gauge inter-participant performance variability, showed significant variation amongst scenarios. The scenario with several emergencies had the most variability (CV = 61.6 %), whereas the scenario with medical emergencies had the lowest variability (CV = 34.7%). This implies that in complicated, multi-emergency scenarios, participant performance fluctuated most greatly, but in medical crises, participant performance remained stable.
- Trial observations and performance metric analysis revealed that subjects occasionally attributed precedence to medical issues and tasks over aviation-related circumstances, even when the latter happened to have a detrimental impact on flight safety as revealed by the lower Coefficient of Variation (figure 3a) and the higher emergency response time for medical emergency scenario (figure 3b)



(a) Inter-participant Variability



(b) Emergency Response Time

Figure 3: Participant behavior

The exploratory data analysis illustrates that although medical professionals with minimal AI and aviation experience are able to interact with an AI pilot to carry out both medical and aviation activities and navigate through emergencies, it appears to be inappropriate to expect them to attribute precedence to aircraft tasks even during aviation-related emergencies. Additionally, these early findings highlight the significance of recognizing the constraints of human-AI collaboration and taking human aspects into account when designing autonomous rotorcraft systems, especially for missions requiring the presence of humans on board, such as medical evacuation missions where simultaneous management of aviation and medical activities is required. The significance of ongoing investigation and advancement in this sector is highlighted by these findings, which might have significant relevance for lawmakers, operators, and manufacturers of rotorcraft. Incorporating human factors concerns is essential for effective and safe operations as we move towards more autonomous technologies. As we continue to analyze, we hope to gain a more nuanced understanding of human-AI interaction.

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. N. Puca and G. Guglieri, “Enabling civil single-pilot operations: A state-of-the-art review,” *Aerotecnica Missili & Spazio*, 2024.

2. D. P. Comerford, S. L. Brandt, J. Lachter, S.-C. Wu, R. Mogford, V. Battiste, and W. W. Johnson, “NASA’s Single-Pilot Operations Technical Interchange Meeting: Proceedings and Findings,” 2013.

3. Y. Lim, V. Bassien-Capsa, S. Ramasamy, J. Liu, and R. Sabatini, “Commercial airline single-pilot operations: System design and pathways to certification,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 32, no. 7, pp. 4–21, 2017.

4. K.-P. L. Vu, J. Lachter, V. Battiste, and T. Z. Strybel, “Single Pilot Operations in Domestic Commercial Aviation,” *Human Factors*, vol. 60, no. 6, pp. 755–762, 2018. PMID: 30063410.

5. M. A. Piera, J. L. Muñoz, D. Gil, G. Martin, and J. Manzano, “A Socio-Technical Simulation Model for the Design of the Future Single Pilot Cockpit: An Opportunity to Improve Pilot Performance,” *IEEE Access*, vol. 10, pp. 22330–22343, 2022.

6. R. Bennett and R. Vijaygopal, “Air passenger attitudes towards pilotless aircraft,” *Research in Transportation Business and Management*, vol. 41, p. 100656, 2021.

7. S. Jones, “Autonomous helicopter technology wins another major award - composites opportunities in eVTOL,” Mar. 2022.

8. J. HARPER, “Autonomous helicopters seen as wave of the future,” *National Defense*, vol. 102, no. 771, pp. 26–27, 2018.

9. R. Agbeyibor, V. Ruia, J. Kolb, and K. M. Feigh, “Joint Intelligence, Surveillance, and Reconnaissance Mission Collaboration with Autonomous Pilots,” *Human Factors and Ergonomics Society Annual Meeting*, Aug. 2024.

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