# Towards Safe Collaboration Between Autonomous Pilots and Human Crews for Intelligence, Surveillance, and Reconnaissance


Richard Agbeyibor
*Georgia Institute of Technology*
richard.agbeyibor@gatech.edu

Vedant Ruia
*Georgia Institute of Technology*
vruia6@gatech.edu

Jack Kolb
*Georgia Institute of Technology*
kolb@gatech.edu

Carmen Jimenez Cortes
*Georgia Institute of Technology*
ccortes9@gatech.edu

Samuel Coogan
*School of Electrical and Computer Engineering*
*Georgia Institute of Technology*
sam.coogan@gatech.edu,

Karen M. Feigh
*School of Aerospace Engineering*
*Georgia Institute of Technology*
karen.feigh@gatech.edu



***Abstract*—Many aviation missions today are accomplished by a heterogeneous crew of pilots and mission specialists. As fully Automated Pilots (AP) are integrated into aviation crews, effective teaming will be necessary for safety assurance and mission effectiveness. This flight simulator study explored teaming between a non-pilot human operator and an AP collaborating on a maritime Intelligence, Surveillance, and Reconnaissance (ISR) mission. The study compared a *Waypoint* AP behavior, requiring human intervention in aircraft control to prevent overflight of damage-causing enemy ships, with a *Collision Avoidance* behavior where the AP proactively avoids enemy ships using control barrier functions. This proactive AP behavior resulted in less aircraft damage and more predictable team performance, albeit longer mission times. Results indicate that situation awareness varied with AP complexity level and task load level. Participants perceived positively the AP when it succeeded and calibrated their trust when it failed.**



***Index Terms*—autonomy, automated pilot, collaboration, teaming, human factors, ISR, control barrier functions**


## I. Introduction

The rapid pace of autonomy development is trending towards many basic aviation functions being competently handled by fully automated pilots (APs) [1] [15] [7]. The history of aviation automation [17], however, shows us that increased complexity can lead to decreased situation awareness (SA), trust, and perception [6]. As we integrate APs into aviation crews, it is therefore necessary to carefully consider human factors to ensure safety and mission effectiveness.

Intelligence, Surveillance, and Reconnaissance (ISR) is the task of persistent monitoring of a target or area. Commonly applied to border and coastline surveillance, military operations, and policing regions of interest, aerial ISR operators typically pilot an aircraft and observe potential threat actors to assess and communicate risk. Opportunities exist to leverage research topics from human factors and controls to automate ISR pilots, and improve the safety and efficiency of ISR missions.

Significant research has addressed autonomous aircraft navigation for mission-based objectives. However, controllers are typically designed for fully-autonomous control without human input nor shared control. In ISR, the human operator is an integral part of a human-AP team, and the AP must be flexible to the operator's desires, human factors limitations, and changes in mission objectives. Little work has explored joint human-AP control in ISR missions, nor how an AP can be considerate of how its flight behavior affects the operator.

In this work we explore how AP complexity affects a human-AP team's mission effectiveness and the operator's human factors. We utilize an immersive ISR mission that tasks a human operator and an AP team with identifying, classifying and tracking ships in an assigned surveillance area. High-threat enemy ships have a weapons engagement zone (WEZ) which damages the ISR aircraft if it enters the WEZ boundaries.

We compare how two AP flight behaviors – *Waypoint* and *Collision Avoidance* – affect the team's mission effectiveness and the human operator's situation awareness and perception. The *Waypoint* AP follows a fixed search pattern, with the operator being responsible for navigating the aircraft around WEZs of enemy ships. The *Collision Avoidance* AP actively avoids the WEZs of enemy ships with the use of planning algorithms and control barrier functions (CBFs) [2, 3, 11, 18].

We hypothesize the following:

- **H1** The *Collision Avoidance* AP will have a higher mission effectiveness than the *Waypoint* AP.
- **H2** The *Collision Avoidance* AP will **not** result in a loss of operator situation awareness when compared to the *Waypoint* AP.
- **H3** Participants will report a more positive perception of the *Collision Avoidance* AP compared to the *Waypoint* AP.

The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Air Force, the Navy, the Department of Defense or the U.S. Government

## II. Background

Many aviation missions today require a heterogeneous crew of pilots and mission specialists. Example civilian missions include disaster relief, air ambulance, and law enforcement. Military examples include search and rescue, medical evacuation, and intelligence, surveillance and reconnaissance (ISR). In the near future non-pilot human crew mates could be asked to team with APs to accomplish these same missions [4].

Human factors and aviation psychology research over the last 30 years have shown that humans are significantly challenged when asked to monitor complex automation [10]. When asked to work with even well designed complex automation or autonomy, people can suffer from poor understanding, poor situation awareness (SA), cognitive biases and inappropriate trust of the system [10].

One of the potential negative consequences of autonomy is a decrement in operator performance, particularly when it comes to situation awareness [6]. Situation awareness (SA) is defined as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [5]. A study of levels of automation in [6] found that full automation produced more problems than did partial automation, which in turn produced more problems than manual operation. As autonomy replaces automation, a similar effect is expected with the increased complexity of autonomy in APs.

Although people hold a generally positive attitude toward collaboration with AI teammates, studies show that they have mixed feelings toward individual AI agents [20]. In contrast to using AI as a tool, treating AI as a teammate means that people put higher expectations on their capabilities [10] [20].

Runtime assurance or safety assurance mechanisms are algorithms that guarantee the safety of intelligent control systems by monitoring the state of the system and intervening when necessary [8]. For safety-critical systems such as aircraft, safety assurance mechanisms prevent conditions that would lead to loss of control, physical damage to the aircraft, loss of human life, or failure of the mission [8].

The integration of advanced autonomous systems into team and crew structures will require new paradigms in human factors to provide safety assurance in increasingly complex safety-critical aviation systems.

## III. Methods

This paper presents the results of an empirical human-AI collaboration study between non-pilot human crewmates and an AP to accomplish an ISR mission. In a flight simulator cabin shown in Fig. 1, participants simulated the role of an intelligence analyst surveying a coastline for armed enemy ships hiding amongst unarmed fishing and cargo ships.

In the scenario, the human analyst collaborates with an AP on control of the ISR aircraft to enable effective classification of the ships in the surveillance area. On the graphical user interface shown in Fig. 2, herein called the ISR Operator Control Station, target ships are characterized by a Weapon Employment Zone (WEZ). The team is tasked to fly the aircraft within sensor range of the target ship, while avoiding overflight of enemy ship WEZs.

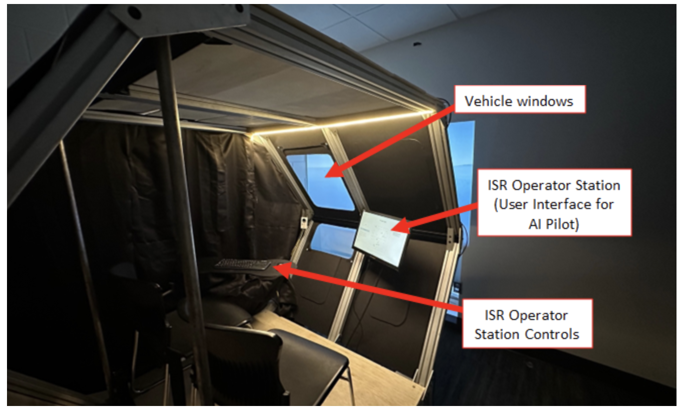


Fig. 1: Simulator Cabin

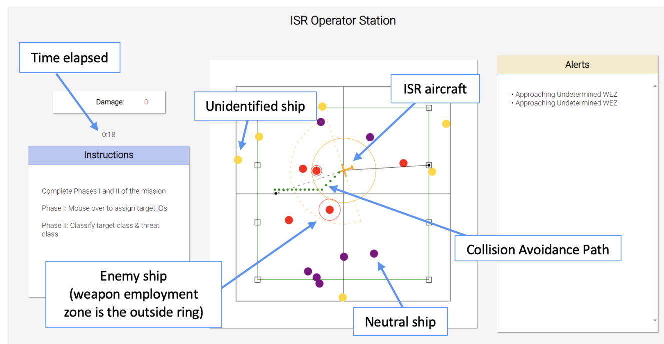


Fig. 2: ISR Operator Station in Collision Avoidance AI Behavior

The independent variables were the complexity level of the AP behavior (2 levels), and the task load level as operationalized by the number of targets to surveil (2 levels). The dependent variables were the SA of the human crew member, crew eye gaze, the perception of the crew of the AP, and the mission effectiveness.

### A. Fully Automated Pilot

The role of the AP is to aviate, navigate, and communicate, while the human analyst classifies the ships. Various behaviors based on AI techniques are tested for the AP. The base AP behavior, 'Waypoint', navigates the surveillance area by flying a pre-briefed search pattern with waypoints marked on the moving map display. The human operator can at any time override the next search pattern waypoint by setting an alternate waypoint for the AP to fly. The second level of AP behavior increases autonomous decision-making by implementing a collision avoidance mechanism using the A* pathfinding algorithm and CBFs. The Collision Avoidance behavior provides a safety mechanism where the AP proactively avoids the WEZ of enemy ships.

### B. Safety Assurance through CBFs

The autonomy's safety mechanism is modeled as a set of constraints in the system's state. CBFs enforce forward invariance of the constraint set so that no trajectory initialized within the constraint set ever leaves or violates the constraint set. In this work, we used CBFs to avoid enemy WEZs. Modeling our autonomous aircraft as a control affine system allows us to include these CBFs in convex optimization programs and obtain a control input that will render the constraint set forward invariant.

Consider the affine controlled system

$$\dot{x} = f(x) + g(x)u, \qquad x \in \mathbb{R}^n,\ u \in \mathbb{R}^m \tag{1}$$

and suppose that the constraint set is closed and defined as $S = \{x \mid h(x) \geq 0\}$, with boundary $\partial S = \{x \mid h(x) = 0\}$, for a continuously differentiable function $h(x)$. $h : \mathbb{R}^n \to \mathbb{R}$ also satisfies that $h(x) = 0$ implies $\nabla h(x) \neq 0$. The set $S$ is forward invariant for the system (1), if for all $T > 0$, all $x_0 \in S$, and all solutions $x(t)$ on $[0, T]$ satisfying $x(0) = x_0$, it holds that $x(t) \in S$ for all $t \in [0, T]$. If, further, $f$ is Lipschitz continuous, it holds that

$$\nabla h(x)^T (f(x) + g(x)u) \geq -\alpha(h(x)) \quad \text{for all } x \in \mathbb{R}^n \tag{2}$$

for some locally Lipschitz function $\alpha : \mathbb{R} \to \mathbb{R}$ satisfying $\alpha(0) = 0$. With this condition, our goal is now to design a feedback controller $u = \sigma(x)$ such that $S$ is forward invariant. Condition (2) leads to the design criterion that any Lipschitz continuous feedback controller $\sigma(x) \in U(x)$ where

$$U(x) = \{u \mid \nabla h(x)^T (f(x) + g(x)u) \geq -\alpha(h(x))\} \tag{3}$$

ensures forward invariance of $S$. Notably, the inequality in (3) is affine in $u$ and, therefore, can be included in convex optimization programs to compute a feedback controller $\sigma(x)$ at runtime. If such a feedback controller exists, then $h(x)$ is called a CBF.

One possible approach to obtain this new *safe* controller is to solve the Quadratic Program (QP) in (4)

$$\begin{aligned} &\underset{u}{\text{minimize}} \quad ||u - \hat{u}||^2 \\ &s.t. \quad \dot{h}_i \geq -\alpha_i(h_i) \quad \forall i = 1, ..., N \end{aligned} \tag{4}$$

where $N$ is the total number of enemy targets.

The AP behavior named Collision Avoidance implements a safety assurance mechanism using CBFs around each enemy's WEZ. These CBFs avoid overflight of enemy ship WEZs. Success requires that the CBFs be sufficiently apart so that no more than one CBF constraint will be active at a time such that a solution for (4) exists. In practice, this requirement will be highly dependent on the task load level and the trajectory each enemy target ship follows. We consider this collaborative autonomy behavior an increase in autonomy from the base reactive autonomy Waypoint behavior.

### C. Situation Awareness

Related literature shows that humans are poor supervisors of automation, so increased complexity often leads to decreased SA. We evaluated SA after each scenario through questions that assessed the user's understanding of pertinent information such as the approximate number of enemy ships, the engagement range of ships based on their WEZs, and the distribution of ships in the surveillance area. Examples of SA questions asked include:

- **Low Task Load *Waypoint***: "How many purple targets appeared?". Participants were asked to type in the approximate number into a text box. The question was graded pass/fail with a leniency range of $\pm 1$ for the correct answer.
- **High Task Load *Waypoint***: "How many red targets appeared?"
- **Low Task Load *Collision Avoidance***: "How many red boats had large WEZs / rings around them?"
- **High Task Load *Collision Avoidance***: "What quadrant of the search area was least populated?". Participants selected from a drop down list.

### D. Perception of the AP & Trust

Following the simulation, participants were given a questionnaire and a debrief interview with various questions about their perception of the AP.

The questionnaire was administered following the last scenario while participants were still seated in the aircraft cabin. The questions were sourced from the Interdependent Trust for Humans and Automation Survey (I-THAu) [12]. The questions targeted affective, general and structural trust [13].

The debrief interviews were semi-structured with five questions that assessed positive characteristics of the AP, negative characteristics, user trust and user perception of the AP. The last question asked for any other feedback or comment they wanted to share about the AP or their experience.

1) What did you like about the AP?
2) What did you not like about the AP?
3) Did you feel like you could trust the AP in real life?
4) Did you feel like the AP was beneficial to you and your mission goals in this scenario?
5) Do you have any additional comments or feedback about working with the AP?

### E. Mission Effectiveness

Mission effectiveness was assessed by the participant's damage and time to complete each trial. A composite score was calculated using Equations 5 and 6.

$$Score_{user} = D + T \tag{5}$$

Where $D$ is the damage taken during the trial, and $T$ is the time to complete the trial. For our evaluation, we normalized each term by the maximum damage taken $D_{max}$ and the maximum time to complete $T_{max}$ over all users, and negated and scaled the score. The resulting normalized score is scaled between $[0, 1]$, where *higher is better*.

$$Score_{norm} = \frac{2 - (\frac{D}{D_{max}} + \frac{T}{T_{max}})}{2} \quad (6)$$

## IV. User Study Design

Twenty-eight participants were recruited from our university and local area. Participant ages ranged from 19 to 35 years old with a mean age of 23. 70% of participants were male, 22% were female, and 8% were non-binary. The study required 2.5 hours of a participant's time, and participants were compensated \$50. The study protocol was approved by the university institutional review board.

None of the participants had prior ISR experience. Six participants had some AI experience and one had prior flight experience. They received basic training on the AP, the ISR mission and the aircraft dynamics. This demographic is representative of new recruit intelligence analysts who may be thrust into a real world mission with minimal training due to world events.

## V. Results

### A. Mission Effectiveness

Adding the safety assurance mechanism in Collision Avoidance resulted in lower damage, albeit an increased mission duration. Fig. 3 shows the normalized score (Eq. 6) for each AI behavior in low and high task load conditions. Notably, the Collision Avoidance behavior decreased the interquartile range (IQR) of participant's mission effectiveness. As the AP exercised more control over the flight trajectories, the overall team's mission effectiveness was more predictable.

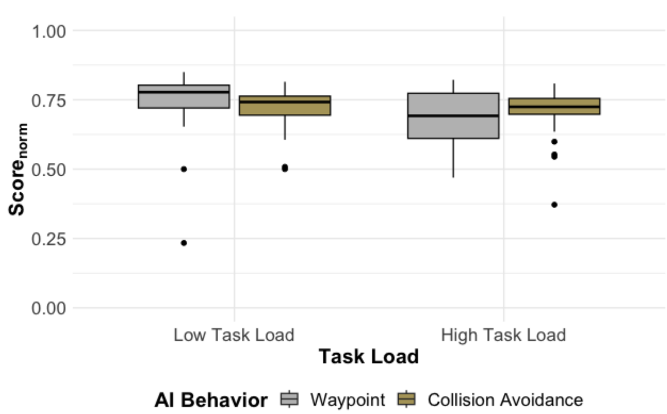


Fig. 3: $Score_{norm}$ vs. AI Behavior

Fig. 4 shows the damage scores for the Waypoint and Collision Avoidance behaviors across task loads. In low task load conditions, participants accumulated negligible damage with both AP behaviors. In the high task load scenario, the increase in the number and speed of enemy ships led to an increase in damage score with the Waypoint behavior. The Collision Avoidance behavior effectively assured safety and minimized damage even in high task load conditions.

In a few instances, the QP in (4) failed to find a solution that met the constraint set for the high number of ships in the

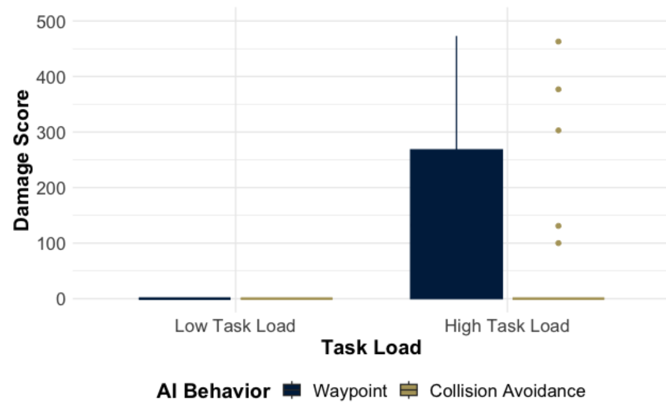


Fig. 4: Damage Score vs. AI Behavior

high task load scenario. As mentioned, there are no formal guarantees on finding a solution to the QP in (4) if there are overlapping WEZs or if their trajectories are such that there is no available danger-free path for the aircraft to follow and the enemies continue to get closer to the aircraft.

In the damage score box plot for Collision Avoidance under high task load shown in Fig. 4, five data points that are more than two standard deviations from the mean are not plotted. The five damage score data points that occured with Collision Avoidance under high task load were due to failures of the CBFs.

When the CBFs failed in the Collision Avoidance behavior, participants incurred damage if they did not manually reroute the aircraft as it flew towards and into an enemy WEZ. This an example of over reliance of the crew on the AP and a failure to over-ride its behavior when the human needed to jump back in the loop [19].

### B. Perception of the AP

The answers to the post-trial questions about the AP are shown in Fig. 5. Although most aspects of the AP were perceived neutrally, participants generally viewed the AP's collision avoidance positively, recognizing the feature as beneficial to their mission goals.

*1) Debrief Interviews:* The debrief interviews assessed positive characteristics of the AP, negative characteristics, user trust, user perception, and general feedback on the system. The debrief interviews ranged from 3-14 minutes in length with some participants having strong opinions and some being very ambivalent.

Generally, participants liked the assistance with navigation provided by the Waypoint behavior and the ease of mind provided by the Collision Avoidance Behavior.

Participants did not like the lack of transparency of the system, and the slow response of the AP to user inputs. The majority of participants said that in real life, they would be reluctant to trust the AP with precise navigation in close proximity to enemy ships but were willing to trust it with general navigation and maneuvering of the aircraft in non-critical situations.

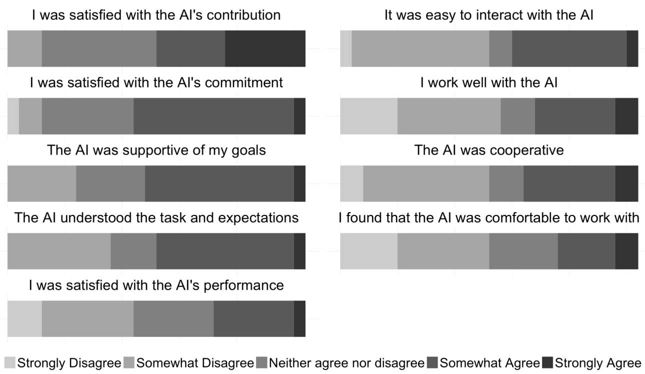


Fig. 5: Post-Trial Questionnaire Responses

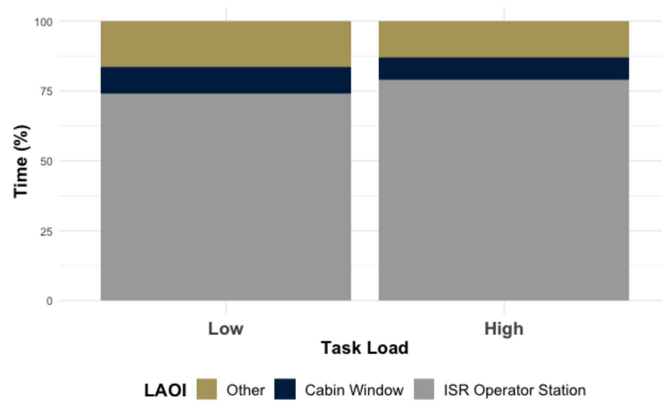


Fig. 6: Participant Gaze in Areas of Interest

Participants were mostly neutral on the benefit of the AP to the effectiveness of their mission. Most expressed that the mission took longer because the AP took a wide turning radius, maneuvered sluggishly, and was not flexible in its search pattern. Many expressed that the Collision Avoidance behavior enabled them to delegate navigation while accomplishing the human only classification tasks.

A few participants had strong preferences for choosing the search pattern themselves and continuously overrode the AP's default search pattern. This study did not measure pre-dispositional attitude so it is not possible to investigate whether there was a shift in the intensity of participant perception due to the AP behaviors.

A couple participants attributed intent to the AP's behavior and interpreted its actions as deliberate steps to accomplish its own goals or to communicate suggestions to the human operator. These two participants expressed that they were unfamiliar with AI technology and had come in to the experiment with expectations of a highly intelligent agent with sophisticated decision-making and communications skills. The AP system was not designed with such an architecture, and so reality did not meet their expectations.

The occasional failure of the CBFs in the Collision Avoidance behavior negatively impacted participants' perception of the system. Participants indicated that this decline in expected capability of the AP decreased their trust in the system.

*2) Questionnaires:* Participants were asked to complete a post-experiment questionnaire that was aimed at gauging their perspective on working with the AI in the AP. Participants indicated their level of agreement to 11 statements. The results of the post-experiment questionnaires are shown in Fig. 5. Along the horizontal axis, darker shades represent agreement to the statement while lighter shades show disagreement as indicated in the legend.

Notable findings show that users were satisfied with the AI's teammate characteristics of commitment and contribution. Conversely, a majority of users felt that the AI was not comfortable to work with, not cooperative, and not easy to interact with.

### C. *User Experience*

The participants wore Argus Science ETVision eye trackers throughout the experiment which measured their gaze location in the cabin and on the ISR Operator Station. Live Areas of Interests (LAOIs) identified whether the participants were looking outside through the vehicle windows, inside at the control station, or another area of the cabin shown in Fig. 1. For the times when they were looking at the ISR operator control station, heat maps were created for the participant's gaze trajectory. No significant trends were found in the aggregated heat maps across behaviors and task loads.

User gaze location was tracked through three areas of interest: the ISR Operator Station, Cabin Window, and Other. With an increase in task load users spent more time looking at the ISR operator station ($t = -2.2868$, $df = 214$, p-value $= 0.02319$) and less time looking elsewhere as shown in Fig. 6. When comparing users' gaze location across AP behaviors, there was no significant change; users looked at the ISR operator's station in the Collision Avoidance mode just as much, even with its safety assurance mechanism.

### D. *Situation Awareness*

When looking at specific task load levels, as shown in Fig. 7, results show that SA did not change across behaviors in the low task load conditions. In the high task load conditions, SA decreased with increased autonomy complexity where 70.3% of participants passed the SA question in Waypoint and only 44.4% passed in the Collision Avoidance scenario.

## VI. Analysis and Discussion

The fully Automated Pilot technology reduces the amount of damage that users accrue throughout each scenario as seen in Fig. 4. The primary benefit of the Collision Avoidance behavior is observed in the high task load scenario. This shows that there is real value-add of a safety assurance mechanism when the team is highly tasked such that the human analyst does not have sufficient excess cognitive capacity to closely monitor the AP. Although times to complete the mission were longer with increased automation complexity, the safety

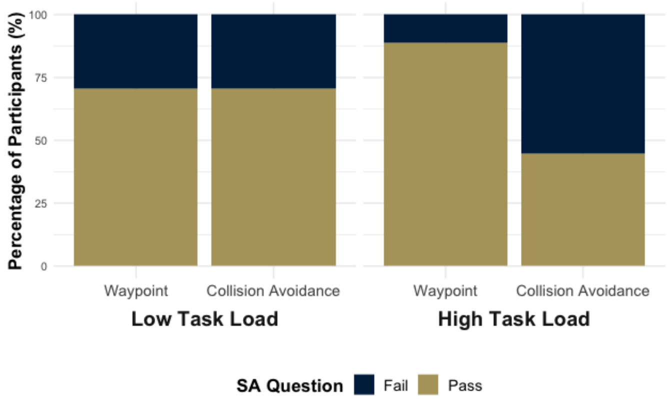


Fig. 7: Situation Awareness vs. AI Behavior per Task Load

benefits may outweigh the time duration costs depending on the situation. When given equal weighing, as shown in the normalized performance scores in Fig. 3 they did outweigh. Crew may initially be reluctant to give up navigation control to an AP due to its slower mission completion, however, results show that advanced AI-based APs can be a hallmark for safe collaboration between automated pilots and human crews. These results affirm hypothesis **H1**.

There are few outliers in the damage score due to failures of the control barrier functions. The AP in this study did not warn the user when its CBF QP solver failed. It just continued to fly towards and into enemy WEZs. Detection of an impending failure required vigilance on the part of the analyst. The damage incurred in this study are an indication of the dangers of over reliance on automation and autonomy and the seeming randomness of failures for complex and highly reliable automation. Participants that suffered failures of the CBF did indicate in their debrief interviews that their trust fell after a failure. This led them to calibrate their trust more appropriately. Participants expressed that they were comfortable trusting the AP with general navigation in open areas but would not trust it in real life with precise navigation in close proximity to enemy ships.

Table I shows the interquartile range (IQR) of the normalized score for the various task load and AP behavior conditions. Due to a few extreme outliers that skew the variance, IQR is used to show the range of scores that the bulk of participants achieved.

In the low task load conditions, participants achieved relatively high scores with both AP behaviors. Increasing the task load conditions to high resulted in a lower score on average and a greater IQR for the base Waypoint behavior. The Collision Avoidance behavior, on the other hand, resulted in a smaller IQR in high task load conditions. As such, an observed advantage of increased automation is a narrower spread in performance and more predictability of outcomes, particularly in high task load conditions. In a real world military scenario where new recruits could fill the role of the human analyst on a similar ISR mission, a small IQR assures a minimum expect-able quality even from the least skilled operator.

Table I: Variances in User Performance Score

| Task Load | Behavior | $Score_{norm}$ Interquartile Range |
|---|---|---|
| Low | Waypoint | 0.083 |
| Low | Collision Avoidance | 0.069 |
| High | Waypoint | 0.163 |
| High | Collision Avoidance | 0.057 |

In real-world operations, safety is the ultimate priority on manned aircraft, but in certain cases minimizing mission duration might take priority over safety of the vehicle if the operation is unmanned. In this study, equal weighting is given to mission duration and damage score (as shown in Eq. 5 and 6). The mission of the vehicle and whether it is manned or unmanned should serve as guidance for the amount of automation to be integrated.

In this experiment the Collision Avoidance AP behavior showed on one hand, the safety benefit of more predictable and consistent performance. On the other hand, it also showed that operators could be less engaged and unable to notice or recover from a failure. These insights reveal that with continued advances in autonomy and automation technology, if humans fall out of the loop of a system and lose situation awareness, system failures will likely be more detrimental [6]. Conversely, this experiment also supports that automation does not always mean operational efficiency when it comes to working with humans.

In the low task load scenario, user situation awareness did not change across AP behavior. In the high task load scenario, user situation awareness dropped with an increase in automation (Fig. 7). These results affirm hypothesis **H2** in low task load conditions, and refute it in high task load conditions. The observed decrease in situational awareness in the high task load conditions could be explained by participant over reliance on the system; during the post-experiment interviews several users expressed appreciation for the comfort and assistance provided by the Collision Avoidance behavior. The observed decrease in situational awareness could also be explained by a task load that exceeded participants' capacity to multi-task to accomplish both their analyst classification task and AP monitoring task. Participants in their debriefs validated the analysis of the results shown in Fig. 7, that the Collision Avoidance AP was most useful during the high task load conditions.

In real-world systems 100% runtime assurance is rarely guaranteed. Literature shows that a drop in SA compromises the safe operation of automated systems. In the automotive sector, companies have taken the approach of utilizing engagement prompts and verification mechanisms to keep the user engaged in order to mitigate reduction in SA [16] [9]. These operator engagement prompts [14] could be simple like requiring frequent user input on controls such as a steering wheel or it could be more complex such as monitoring the gaze location of user's eyes.

Notably in our research, eye tracking results indicated that even though users spent more time looking at the ISR Operator

Station in high task load conditions, their SA did not improve. Future work for this research involves the addition of an adaptive AP that utilizes human engagement data obtained through physiological sensors to modulate the speed and flying characteristics of the aircraft. This approach to safety assurance benefits mission effectiveness by minimizing the effects of a lack of situation awareness in the operator, rather than trying to forcibly raise their SA.

The questionnaire results showed more positive answers to questions addressing AI task work such as “I was satisfied with the AI’s contribution” shown in the left column of Fig. 5. The results showed more negative answers to questions addressing the AI’s team work such as “It was easy to interact with the AI” shown in the right column of Fig. 5. In concert with the questionnaire results, participants’ expressed appreciation for the Collision Avoidance affirms hypothesis **H3**.

In addition to creating safety assurance features, system designers should invest in improving the interaction characteristics of their APs. Improvements can range from enhanced interaction through different modalities such as audio input and feedback to an increase in AI explain-ability and transparency. The primary complaint that users expressed in the post-experiment debrief was a lack of system transparency; several participants mentioned that the AP would not initiate commanded turns to points immediately. Immediate turns are not possible due to vehicle and controller dynamics. Mapping the curved path of the vehicle is a viable method of providing users more usable information compared to the current overly-simplified straight-line depiction.

## VII. Conclusion

In this work we explore how the flight behavior of an AP in a human-AP ISR team affects the mission effectiveness and the human operator’s perception and situation awareness. We introduce a novel application of CBFs in a coupled human-autonomy teaming domain. Through a user study, we find that participants preferred the CBF-based *Collision Avoidance* behavior, and appropriately calibrated their trust to failures with the AP behavior.

In line with the literature, the increased level of autonomy resulted in decreased situation awareness in a high task load scenario, motivating the importance of considering the effects of high-autonomy systems on operator situation awareness.

Our work has several limitations that can be addressed in future work. While the high task load scenario was sufficiently complex to decrease the operator’s situation awareness with the *Collision Avoidance* AP, we did not find a meaningful decrease in the *Waypoint* AP. Future work can increase the complexity of the ISR domain to enable a convincing comparison between AP behaviors.

Additionally, due to the high number of enemy targets and their unpredictable trajectories in the high task load scenario, the CBF runtime assurance mechanism failed to find a solution on some occasions. Future work could explore further assumptions in modelling of the constraints to provide some formal guarantees of the performance of CBFs and improve the reliability of the AP.

Furthermore, our study was conducted with participants with no ISR experience. Expert operators would have had different baselines for task load and situation awareness due to their familiarity with the mission. Trust and perception of the benefit of the AI would also be affected by expertise and experience of the human crewmate.

Lastly, our ISR simulator used a simplified aircraft model with unrealistic flight dynamics. We are interested in integrating a higher degree of freedom model to better align with real-world systems.

This work lays foundations for the development of AI pilots or AI enabled APs in future autonomous aircraft. The latest news from the Department of Defense indicate significant investments in development of uncrewed autonomous vehicles. This research looks beyond that to crewed autonomous vehicles and the human factors necessary for safe collaboration between onboard human operators and their AI pilot crewmates.

Our results indicate that is possible to reduce crew requirements on missions like ISR that currently require a human crew of pilots and non-pilot operators. In addition to reduction in manpower costs, our results indicate that it is possible to assure a predictable range of safety and mission effectiveness even with the least trained operator using run time assurance mechanisms like that of our Collision Avoidance AP. Maturation of systems like this will provide for safe collaboration between autonomous pilots and human crews.

## Acknowledgment

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 Air Force, the Navy, the Department of Defense or the U.S. Government.

## References


[1] *Air Combat Evolution*. URL: https://www.darpa.mil/program/air-combat-evolution (visited on 03/04/2024).

[2] Aaron D Ames, Jessy W Grizzle, and Paulo Tabuada. "Control barrier function based quadratic programs with application to adaptive cruise control". In: *53rd IEEE Conference on Decision and Control*. IEEE. 2014, pp. 6271–6278.

[3] Aaron D Ames et al. "Control barrier functions: Theory and applications". In: *2019 18th European control conference (ECC)*. IEEE. 2019, pp. 3420–3431.

[4] "Collaborative Combat Aircraft (CCA), USA". In: *Airforce Technology* (). URL: https://www.airforce-technology.com/projects/collaborative-combat-aircraft-cca-usa/ (visited on 03/12/2024).

[5] Mica R. Endsley. "Design and Evaluation for Situation Awareness Enhancement". In: *Proceedings of the Human Factors Society Annual Meeting* 32.2 (Oct. 1, 1988). Publisher: SAGE Publications, pp. 97–101. ISSN: 0163-5182. DOI: 10.1177/154193128803200221. URL: https://doi.org/10.1177/154193128803200221 (visited on 04/03/2024).

[6] Mica R. Endsley and Esin O. Kiris. "The Out-of-the-Loop Performance Problem and Level of Control in Automation". In: *Human Factors* 37.2 (June 1, 1995). Publisher: SAGE Publications Inc, pp. 381–394. ISSN: 0018-7208. DOI: 10.1518/001872095779064555. URL: https://doi.org/10.1518/001872095779064555 (visited on 01/21/2024).

[7] D. Harris. "Single-pilot airline operations: Designing the aircraft may be the easy part". In: *The Aeronautical Journal* 127.1313 (July 2023), pp. 1171–1191. ISSN: 0001-9240, 2059-6464. DOI: 10.1017/aer.2022.110. URL: https://www.cambridge.org/core/journals/aeronautical-journal/article/singlepilot-airline-operations-designing-the-aircraft-may-be-the-easy-part/A5C6DA432BC5A59EF4973F2F44E5C590 (visited on 03/04/2024).

[8] Kerianne L. Hobbs et al. "Runtime Assurance for Safety-Critical Systems: An Introduction to Safety Filtering Approaches for Complex Control Systems". In: *IEEE Control Systems Magazine* 43.2 (Apr. 2023). Conference Name: IEEE Control Systems Magazine, pp. 28–65. ISSN: 1941-000X. DOI: 10.1109/MCS.2023.3234380. URL: https://ieeexplore.ieee.org/abstract/document/10081233 (visited on 04/03/2024).

[9] Andrew L. Kun. "Human-Machine Interaction for Vehicles: Review and Outlook". In: *Foundations and Trends® in Human–Computer Interaction* 11.4 (Apr. 4, 2018). Publisher: Now Publishers, Inc., pp. 201–293. ISSN: 1551-3955, 1551-3963. DOI: 10.1561/1100000069. URL: https://www.nowpublishers.com/article/Details/HCI-069 (visited on 07/06/2022).

[10] Engineering National Academies of Sciences. *Human-AI Teaming: State-of-the-Art and Research Needs*. Dec. 22, 2021. ISBN: 978-0-309-27017-5. DOI: 10.17226/26355. URL: https://nap.nationalacademies.org/catalog/26355/human-ai-teaming-state-of-the-art-and-research-needs (visited on 09/09/2022).

[11] Stephen Prajna, Ali Jadbabaie, and George J Pappas. "A framework for worst-case and stochastic safety verification using barrier certificates". In: *IEEE Transactions on Automatic Control* 52.8 (2007), pp. 1415–1428.

[12] Yosef Razin. *Interdependent Trust for Humans and Automation Survey*. Unpublished manuscript. 2020.

[13] Yosef S. Razin and Karen M. Feigh. *Converging Measures and an Emergent Model: A Meta-Analysis of Human-Automation Trust Questionnaires*. Mar. 24, 2023. DOI: 10.48550/arXiv.2303.13799. arXiv: 2303.13799[cs]. URL: http://arxiv.org/abs/2303.13799 (visited on 08/07/2023).

[14] Karina Roundtree, Michael A. Goodrich, and Julie A. Adams. *Transparency: Transitioning From Human–Machine Systems to Human-Swarm Systems*. DOI: 10.1177/1555343419842776. URL: https://journals.sagepub.com/doi/epdf/10.1177/1555343419842776 (visited on 01/18/2024).

[15] David Roza. *AFSOC Flies 3 Reapers With One Crew In First-of-Its-Kind Exercise*. Air & Space Forces Magazine. Jan. 26, 2024. URL: https://www.airandspaceforces.com/afsoc-multiple-mq-9-reapers/ (visited on 03/04/2024).

[16] Tesla. *Update Vehicle Firmware to Prevent Driver Misuse of Autosteer — Tesla Support*. Tesla. URL: https://www.tesla.com/support/vehicle-firmware-prevent-autosteer-misuse (visited on 04/26/2024).

[17] Earl L. Wiener and Renwick E. Curry. "Flight-deck automation: promises and problems". In: *Ergonomics* 23.10 (Nov. 1, 1980). Publisher: Taylor & Francis, pp. 995–1011. ISSN: 0014-0139. DOI: 10.1080/00140138008924809. (Visited on 10/28/2022).

[18] Wei Xiao, Christos G Cassandras, and Calin Belta. *Safe Autonomy with Control Barrier Functions: Theory and Applications*. Springer Nature, 2023.

[19] X. Jessie Yang et al. "Evaluating Effects of User Experience and System Transparency on Trust in Automation". In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '17. New York, NY, USA: Association for Computing Machinery, Mar. 6, 2017, pp. 408–416. ISBN: 978-1-4503-4336-7. DOI: 10.1145/2909824.3020230. URL: https://dl.acm.org/doi/10.1145/2909824.3020230 (visited on 02/03/2024).

[20] Rui Zhang et al. ""An Ideal Human": Expectations of AI Teammates in Human-AI Teaming". In: *Proceedings of the ACM on Human-Computer Interaction* 4 (CSCW3 Jan. 5, 2021), 246:1–246:25. DOI: 10.1145/3432945. URL: https://dl.acm.org/doi/10.1145/3432945 (visited on 04/03/2024).