

Joint Intelligence, Surveillance, and Reconnaissance Mission Collaboration with Autonomous Pilots

Richard Agbeyibor, Vedant Ruia, Jack Kolb and Karen M. Feigh

Georgia Institute of Technology

This study explores human-autonomy collaboration between a future autonomous pilot and a human crew member pursuing a joint Intelligence, Surveillance, and Reconnaissance (ISR) mission. We introduce a novel open-sourced autonomous ISR interaction domain simulating real-world scenarios. As aviation increasingly integrates autonomy, our focus lies in understanding how various autonomous capabilities and interface features affect trust, perception, and user interactions. Conducting an exploratory study with 27 participants in a flight simulator, we examine the impact of various autonomy modes on human-autonomy interaction. Through demographic analysis, interface usage, and qualitative responses, we explore the impact of experience, expertise, and authority on collaboration dynamics.

INTRODUCTION

This research is interested in the potential use of autonomy in crewed Intelligence, Surveillance, and Reconnaissance (ISR). It investigates collaboration between a futuristic autonomous pilot and a human crew member performing joint ISR tasks in a human-AI dyad. We present a novel human-autonomy interaction domain that mimics real-world ISR scenarios in a flight simulator and offers an experimental apparatus for researching adaptive teaming and fluency in multitasking and high task load environments. Our work is motivated by the increasing maturity of Artificial Intelligence (AI) technologies and the opportunity to develop collaborative autonomous pilots that can achieve performance on par with or superior to current operational capabilities. The first part of this study involves the design of the modes and behaviors of an autonomous pilot for joint ISR. The second part is an exploratory study conducted in a flight simulator with 27 participants to investigate the effects of autonomy mode and participant background on human-AI interaction.

An active area of research and development is the integration of AI into aviation, particularly through autonomous piloting capabilities (DARPA, 2020). The increasing capabilities of autonomous technologies will enable interaction that closely mimics human peer-to-peer collaboration (Goodrich & Schultz, 2008). A report on human-AI teaming research gaps and priorities by the National Academies of Sciences (2021) encourages leveraging the beneficial dynamics seen in human-human teams for communication and coordination.

Beyond airline operations, many aviation missions today require a heterogeneous crew of pilots and mission specialists such as emergency medical technicians, rescue operators, law enforcement officers, and intelligence analysts. It is possible that in the near future, autonomous pilots will have the maturity to competently handle basic aviation functions currently handled by human pilots. The U.S. military has indicated interest and is investing heavily into research and development of such capabilities (Roza, 2024), (“Collaborative Combat Aircraft (CCA)”, 2023). Although an autonomous pilot could soon “aviate, navigate, and communicate” without human input, certain critical portions of the mission such as medical care, intelligence analysis and other expert tasks would still require a human mission specialist.

In the literature on human-AI and human-autonomy interaction, few researchers have explored interaction in heterogeneous teams performing highly interdependent tasks that are disjunctive, conjunctive or discretionary (Steiner, 1972). ISR is a class of task with such team and task characteristics, and requires specialized mission knowledge, experience, and judgement that typically requires human mission specialists. Although some ISR can be conducted remotely using Remotely Piloted Aircraft (RPA) or Uncrewed Aerial Systems (UAS), many missions require onboard personnel due to their sensitivity, time, sensor, communication, equipment or other operational requirements. As such, our guiding research question is to identify the elements that impact the effective teaming of an autonomous pilot with an onboard ISR crew. This study explores how autonomy mode and operator background impact this interaction.

BACKGROUND

Research in human factors has demonstrated that humans face significant challenges supervising automation (Rasmussen, 1982) (Endsley & Kiris, 1995). Aviation Psychology emerged as a field of study to investigate ways to mitigate those hazardous safety and performance issues in the specific context of aviation (Wiener & Curry, 1980). With the increasing complexity of flight automation and avionics, research has continued into optimizing cockpit displays and autopilot systems (Riley et al., 2002). Much of this research has remained focused on the challenges airline pilots face in their cockpits. As airlines consider single pilot operations (Harris, 2023), additional research will be needed to maintain the safety expected in commercial aviation.

Levels of interaction with automation and autonomy can be classified on a spectrum from direct control to dynamic autonomy based on the capabilities and interface features of the agent, spanning: teleoperation, mediated teleoperation, supervisory control, collaborative control and peer-to-peer collaboration (Goodrich & Schultz, 2008). The increasing capabilities of autonomous technologies are enabling interaction that falls closer and closer to peer-to-peer collaboration (“Collaborative Combat Aircraft (CCA)”, 2023). Summarizing the work of many researchers over the past 50 years on the science of teams, Salas et al. (2008) highlight the performance benefits of dynamic teaming that could be

achieved with sufficiently capable autonomous agents.

In addition to capabilities of the autonomy, the nature of the task is another important characteristic that determines the success of human-AI collaboration. Steiner (1972) established a taxonomy that classifies joint tasks by component, focus, and interdependence. If a task can be divided into clearly identifiable sub-tasks that can be assigned to individual team member, its component categorization is Divisible. On the other hand, if the task requires the group to work together or can only be completed by one individual, its component categorization is Unitary. Tasks that are focused on producing a greater quantity of deliverable are considered Maximizing in focus. On the other hand, tasks that seek to improve quality or achieve an optimal solution, their focus categorization is Optimizing. Interdependence has five categories. Tasks in which individual contributions can be added together for a greater group deliverable are considered Additive. Tasks which require an averaging of individual contributions are considered Compensatory. Tasks which require the team to determine a single solution are considered Disjunctive. Tasks in which the level of deliverable is determined by the most inferior or weakest contribution are considered Conjunctive. And finally, tasks which allow team members to determine the best combination or permutation of individual contributions are Discretionary.

According to *U.S. Air Force Doctrine 2-0 Intelligence* (2023), ISR is “an integrated operations and intelligence activity that synchronizes and integrates the planning and operation of sensors; assets; and processing, exploitation, and dissemination systems in direct support of current and future operations”. Crewed ISR aircraft perform collection operations which involve “tasking and synchronizing ISR sensors, platforms and exploitation resources to characterize the operational environment, adversary activities, and infrastructure, and to target (i.e., find, fix, track, target) entities in the battlespace” (*U.S. Air Force Doctrine 2-0 Intelligence*, 2023). Crewed ISR is a complex task in component, focus, and interdependence that requires close coordination, communication, and collaboration between crew members.

In addition to the nature of the task, other elements of human interaction with AI and autonomy are the capabilities of the autonomy, the nature of the information exchange, the structure of the team, the learning and training adaptation of the team (Goodrich & Schultz, 2008). The nature of the information exchange touches on explainability (Sanneman & Shah, 2020) and bi-directional transparency (J. B. Lyons, 2013) (Chen et al., 2018) which are active areas of research. The structure of the team implicates considerations of responsibility, authority and function allocation (Feigh & Pritchett, 2014). As AI gains the ability to learn online and adapt its task work and teamwork to the environment and team dynamics, elements of adaptive systems will come to bear (Feigh et al., 2012).

Research on human-AI trust has identified age, gender, and participant background as important factors in human-AI interaction (Y. S. Razin & Feigh, 2023). Dispositional trust, faith in technology are elements affected by the human’s background that may impact the interaction. Some factors that may affect human-AI interaction in the specific context of ISR are AI expertise and flight experience.

As autonomy capabilities emerge, some human factors re-

searchers have started evaluating operationally relevant systems in related domains to ISR. Barnes et al. (2021) investigate human-AI interaction in mission planning and dismounted infantry squad operations. J. Lyons et al. (2022) investigate perception of agent appropriateness in air-to-air combat. This research leverages the findings of these research projects to investigate the potential role of autonomous pilots in crewed ISR.

EXPERIMENTAL APPARATUS

We constructed an immersive flight simulator with a cabin the size of an electric Vertical Take Off and Landing (eVTOL) aircraft to provide an operationally realistic environment for future autonomous ISR. The cabin shown in Fig. 1 provides operator seating, front and vertical reference windows, a control station display, and a mouse and keyboard on a swivelling table. Visualization of the aircraft body and outside environment including ships in the water, is programmed using Microsoft Flight Simulator (MSFS) and projected on a floor to ceiling wall using two projectors. The source code used for our experimental apparatus is available at <https://github.com/gt-cec/onr-isr>.

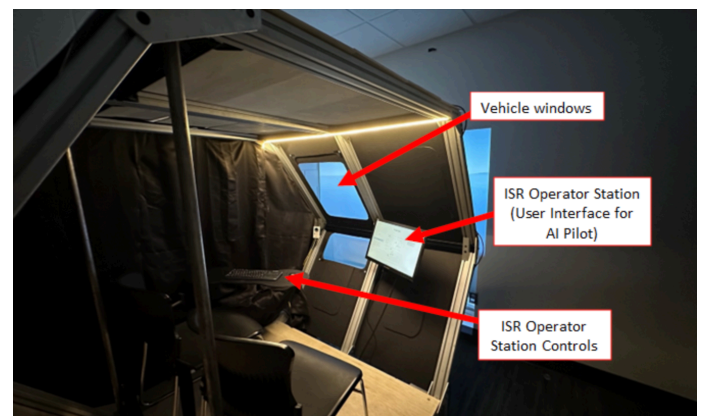


Figure 1: Simulator Cabin

Task Design

In consultation with an operational U.S. Navy P-8 pilot, we designed an ISR Wargame to simulate maritime patrol of a coast line to find adversary ships hiding amidst fishing and cargo ships. The mission is to patrol an assigned surveillance area and fly within sensor range of detected targets to identify, classify, and track adversary ships.

The mission is accomplished in two phases. The first phase – Identify – which can be accomplished from anywhere inside the surveillance area, is to use the ISR operator control station mouse to acknowledge and assign IDs to every target that appears in the surveillance area. The second phase – Classify & Track – first requires navigating the aircraft to within long range sensor range (depicted by the dashed semi-circle around the aircraft in Fig. 2) of a target to classify and color it as a neutral or enemy target. Next, it requires navigating the aircraft within short range sensor range (depicted by the solid circle around the aircraft in Fig. 2) to establish a track and determine the size of the ships’ Weapon Employment Zone (WEZ) if any. The mis-

sion is accomplished when both phases are completed for every target. The task was designed to be complex in component requiring selection and prioritization of targets; in focus requiring both minimization of time and optimization of trajectory; in interdependence requiring selection of a single target to process, averaging the navigation inputs of the human and the AI, and strategic sequencing of targets; and joint.

The team composed of a human intelligence analyst and an autonomous pilot is asked to complete the mission in minimum time and with minimum damage. Damage is accumulated by overflying the WEZ of enemy ships. The team must navigate the aircraft efficiently around the surveillance area to track all targets while taking care to avoid the WEZs of unclassified and enemy ships. The autonomous pilot is tasked with basic aircraft operation and navigation, while the human operator identifies targets and commands the mission to assure completion of all phases. The autonomous pilot is designed with various modes of operation which exhibit different levels of responsibility and authority for aircraft navigation.

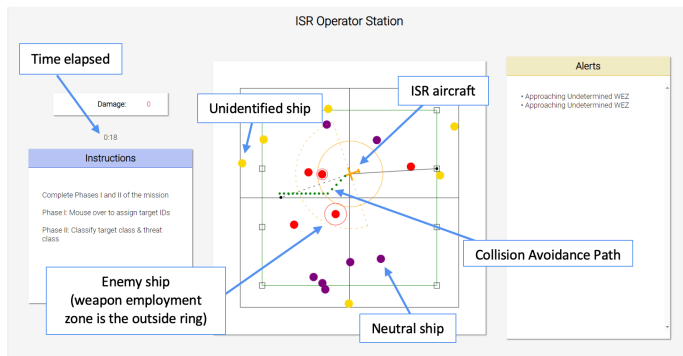


Figure 2: ISR Operator Control Station User Interface

Autonomous Modes

In the baseline mode *Waypoint*, the autonomous pilot defaults to flying a pre-determined navigation pattern but deviates to any arbitrary waypoint requested by the operator. Without any operator input and regardless of the position of enemy ships, the AI flies one of two programmed search patterns: 1) Hold which resembles a rectangular orbit or 2) Ladder which stair steps horizontal scans across the surveillance area. In this mode, the human operator has complete authority over aircraft navigation and responsibility for avoiding enemy WEZs.

The second mode is a dynamics constrained *Collaborative* mode that is different from the baseline mode in that, the autonomous pilot can reject the operator's waypoint request based on a black box maneuver estimation algorithm (designed to mimic maneuvering constraints in safely completing the requested trajectory). The AI does not provide an explanation of its decision. In this mode, the human and the autonomous pilot share authority over navigation, while the human operator retains responsibility for avoiding enemy WEZs.

In *Collision Avoidance* mode, the autonomous pilot continues the default navigation search pattern and also proactively executes evasive maneuvers to avoid overflight of enemy WEZs.

The autonomous pilot accepts all operator waypoints as long as they are not inside a WEZ and navigates to them while executing its evasive maneuvers. The human operator retains authority over aircraft navigation, while sharing responsibility with the autonomous pilot for avoiding enemy WEZs.

In the *Search Optimization* mode, the autonomous pilot has an assistance feature which creates optimized search patterns for navigation. Upon request through a user interface button, the autonomous pilot creates and proposes a new search pattern that focuses navigation in specific quadrants of the surveillance area. The human operator reviews and accepts or declines the new flight pattern. Although the AI can offer assistance, the human operator retains authority over aircraft navigation, and responsibility for avoiding enemy WEZs.

METHOD

Our exploratory user study employed a counter-balanced full-factorial within-subjects design. The independent variables were autonomy mode with four levels and task load with two levels. Each participant completed eight scenarios sequenced according to a Latin Square to counter balance the effects of learning and fatigue. Eight unique Latin Square sequences were sequentially cycled through.

A total of 30 participants were recruited from the university and local community, however, one did not complete training, one did not provide cohesive responses to the questionnaires, and one had incomplete data. 27 participants successfully completed the experiment. The 27 retained participants ranged in age from 19 to 35 with a median of 22 and mean of 22.6. In gender, 19 were male, 6 female and 2 non-binary. For flight experience, 22 had none, and 5 had some including 1 certificated private pilot. For AI experience, 21 had none and 6 had some.

The study protocol was approved by Georgia Tech's Institutional Review Board. The study lasted between 120 and 150 minutes and participants were compensated \$50 for their time. After reviewing the consent forms and completing demographic questionnaires, participants were briefed a background story in which they were a new military intelligence analyst recruit during a cold war in which a recent skirmish destroyed military communications satellites preventing the use of RPAs and UAS for ISR. As such, the military had modified commercial eVTOLs to conduct crewed ISR. Their task was to collaborate with an autonomous pilot to patrol an assigned surveillance area off the coast of the United States. After receiving training on the ISR Wargame and Operator Control Station, participants were instrumented with eye trackers and other physiological sensors. Participants then started completing scenarios according to the Latin Square sequence.

Each scenario lasted between 5 to 10 minutes. Following each scenario, participants completed a NASA TLX questionnaire on their simulator display. After the last scenario and NASA TLX questionnaire, participants completed a comprehensive post-experiment questionnaire on teaming with the autonomous pilot, followed by a debrief interview with one of the researchers.

The post-experiment questionnaire used questions from the Interdependent Trust for Humans and Automation Survey (I-

THAu) (Y. Razin, 2020) to assess team communication, interaction, affective trust and perceived performance. The debrief interview asked participants questions on their positive and negative perceptions of the AI, their trust, and whether they could envision conducting such a collaborative joint human-AI aviation mission in real life. User eye gaze data was collected in addition to mouse movements and clicks. Time and damage score data were recorded for each phase of the mission. System alerts and other user interface messages were also recorded.

RESULTS

In this paper, we employ a mixed methods approach to identify trends in the collected user interface data, the questionnaires, and the debrief interviews for follow-on studies. The objective of our exploratory study was to validate the presence of underlying human factors phenomena and obtain user responses and feedback. We intend to apply user feedback and interview responses from this study to improve our experimental apparatus. We present quantitative findings from our human factors evaluation (workload, situation awareness, physiological metrics) in (Agbeyibor, Ruia, Jimenez Cortes, et al., 2024), and the technical details of the run-time assurance mechanism of our Collision Avoidance mode in (Agbeyibor, Ruia, Kolb, et al., 2024).

In the debrief interviews, participants were asked about their perception of each autonomous mode. Our thematic analysis resulted in the following high-level perceptions:

Waypoint Mode (baseline): Participants had neutral perceptions of this mode, and desired the ability to configure the search pattern behavior.

Collaborative Mode (AI can reject user waypoints): Participants reported negative perceptions because of the lack of explanation in the AI's decisions.

Collision Avoidance Mode (AI avoids enemy WEZs): Participants had a clear preference for this mode, which left the user with authority on navigation, relegating the AI to a supporting role in the team. Participants stated that the mode reduced their workload by allowing them to focus on the task objectives instead of avoiding WEZs.

Search Optimization Mode (dynamic search patterns): Participants had neutral perceptions of this mode, as it did not actively avoid enemy WEZs.

The post-experiment questionnaires measured communication and interaction, affective trust, and perceived performance. T-test revealed no differences between participant demographic groups for communication and interaction, or affective trust. Fig. 3 shows a summary of the questionnaire results.

Three of the questionnaire results indicated small but not statistically significant differences. Participants without AI experience tended to be less satisfied with the AI's performance. Participants without AI experience also tended to be less satisfied with the AI's contribution. And participants without AI experience tended to disagree more with the statement "I work well with the AI" ($\Delta\mu = 1.05, p = 0.1$). In their debrief interviews participants without AI experience expressed that they came in to the experiment with a higher expectation of the AI's capabilities than they experienced.

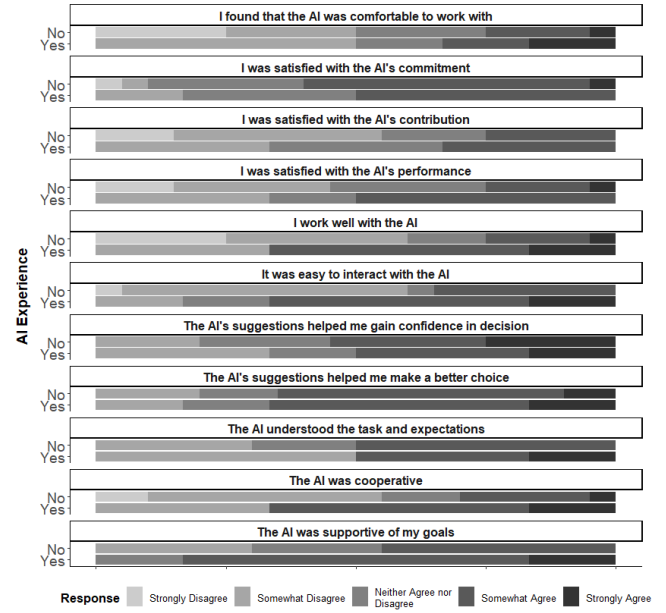


Figure 3: Questionnaire Responses per AI Experience

Mouse click data collected from the user interface suggested a trend in waypoint input differences between participants with flight experience ($\mu = 107.3$) and those without ($\mu = 138.7$), however, the difference was found to not be statistically significant. No statistically significant differences were found in mouse movements or clicks.

DISCUSSION

Overall, the autonomous capabilities provided and the interface features enabled this simplified ISR task to be completed with reasonable performance. The participants' backgrounds indicated differences in perception of the AI pilot, but limited numbers prohibit drawing conclusions. Finally, the perception of the AI was tied to the specific Mode capabilities as expected.

Limitations

One limitation of our work was an imbalance in participant age, gender, AI experience and flight experience demographics, as these were not controlled for. In addition, we did not quantitatively measure differences in autonomous modes beyond the user interface data. The questionnaires did not ask users to rank the various autonomous modes.

In our next iteration, we will target our participant recruitment efforts towards a more even balance of participants across our target demographics. We also intend to measure pre and post-experiment trust to gauge the impact of dispositional trust on attitudes towards the AI. Additionally, we will increase the complexity of the ISR task to strengthen the effect of the autonomous modes on mission outcomes and user experience.

Conclusions

The first takeaway from our preliminary study is that it is possible to conduct an ISR mission with a heterogeneous crew of human

and AI members collaborating on joint aircraft navigation, target identification, classification and tracking.

The guiding question for this paper was how a participant's background and the autonomy's mode of behavior influences the interaction. The research question was addressed through an analysis of relationships between participant demographics, user interface data, and qualitative responses to questions on communication, interaction, trust and perception. Participants expressed a preference for retaining full authority while sharing responsibility with the autonomous pilot for missions tasks. Participant's flight and AI experience played a minor role in the quality of the interaction.

Our future work aims to further develop the autonomous ISR simulator to enable more intelligent navigation pattern optimizations, dynamic adaptation of the autonomous mode to the situational context, and to improve the system's transparency and explainability.

ACKNOWLEDGMENTS

This work was funded in part by the Office of Naval Research, Science of Autonomy grant N00014-21-1-2759. 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

- Agbeyibor, R., Ruia, V., Jimenez Cortes, C., Kolb, J., Coogan, S., & Feigh, K. M. (2024). Run time assurance and human AI fluency in manned autonomous intelligence surveillance and reconnaissance. In *AIAA AVIATION 2024 forum*. American Institute of Aeronautics and Astronautics.
- Agbeyibor, R., Ruia, V., Kolb, J., Jimenez Cortes, C., Coogan, S., & Feigh, K. M. (2024). Towards safe collaboration between autonomous pilots and human crews for intelligence, surveillance, and reconnaissance. In *2024 IEEE/AIAA 43rd digital avionics systems conference (DASC)*.
- Barnes, M. J., Wang, N., Pynadath, D. V., & Chen, J. Y. C. (2021). Chapter 10 - human-agent bidirectional transparency. In C. S. Nam & J. B. Lyons (Eds.), *Trust in human-robot interaction* (pp. 209–232). Academic Press. doi: 10.1016/B978-0-12-819472-0.00010-1
- Chen, J. Y. C., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. , *19*(3), 259–282. doi: 10.1080/1463922X.2017.1315750
- Collaborative combat aircraft (CCA). (2023). Retrieved 2024-03-12, from <https://www.airforce-technology.com/projects/collaborative-combat-aircraft-cca-usa/>
- DARPA. (2020). *Air combat evolution*. Retrieved 2024-03-04, from <https://www.darpa.mil/program/air-combat-evolution>
- Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. , *37*(2), 381–394. (Publisher: SAGE Publications Inc) doi: 10.1518/001872095779064555
- Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a characterization of adaptive systems: A framework for researchers and system designers. , *54*(6), 1008–1024. (Publisher: SAGE Publications Inc) doi: 10.1177/0018720812443983
- Feigh, K. M., & Pritchett, A. R. (2014). Requirements for effective function allocation: A critical review. , *8*(1), 23–32. (Publisher: SAGE Publications) doi: 10.1177/1555343413490945
- Goodrich, M. A., & Schultz, A. C. (2008). Human-robot interaction: A survey. , *1*(3), 203–275. (Publisher: Now Publishers, Inc.) doi: 10.1561/11000000005
- Harris, D. (2023). Single-pilot airline operations: Designing the aircraft may be the easy part. , *127*(1313), 1171–1191. doi: 10.1017/aer.2022.110
- Lyons, J., Highland, P., Bos, N., Lyons, D., Skinner, A., Schnell, T., & Hefron, R. (2022). Measuring perceived agent appropriateness in a live-flight human-autonomy teaming scenario. , 10648046221129393. (Publisher: SAGE Publications Inc) doi: 10.1177/10648046221129393
- Lyons, J. B. (2013). Being transparent about transparency: A model for human-robot interaction.
- National Academies of Sciences, E. (2021). *Human-AI teaming: State-of-the-art and research needs*. doi: 10.17226/26355
- Rasmussen, J. (1982). Human errors. a taxonomy for describing human malfunction in industrial installations. , *4*(2), 311–333. doi: 10.1016/0376-6349(82)90041-4
- Razin, Y. (2020). *Interdependent trust for humans and automation survey*.
- Razin, Y. S., & Feigh, K. M. (2023). *Converging measures and an emergent model: A meta-analysis of human-automation trust questionnaires* (No. arXiv:2303.13799). arXiv. doi: 10.48550/arXiv.2303.13799
- Riley, V., DeMers, B., Misiak, C., & Shackleton, H. (2002). The cockpit control language program: An update or, how i learned to stop worrying and love the CDU. , *III*, 561–566. (Publisher: SAE International)
- Roza, D. (2024). *AFSOC flies 3 reapers with one crew in first-of-its-kind exercise*. Retrieved 2024-03-04, from <https://www.airandspaceforces.com/afsoc-multiple-mq-9-reapers/>
- Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. , *50*(3), 540–547. (Publisher: SAGE Publications Inc) doi: 10.1518/001872008X288457
- Sanneman, L., & Shah, J. A. (2020). A situation awareness-based framework for design and evaluation of explainable AI. In D. Calvaresi, A. Najjar, M. Winikoff, & K. Främling (Eds.), *Explainable, transparent autonomous agents and multi-agent systems* (pp. 94–110). Springer International Publishing. doi: 10.1007/978-3-030-51924-7_6
- Steiner, I. D. (1972). Group process and productivity.
- U.S. air force doctrine 2-0 intelligence. (2023). Retrieved 2024-05-15, from <https://www.doctrine.af.mil/>
- Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: promises and problems. , *23*(10), 995–1011. (Publisher: Taylor & Francis) doi: 10.1080/00140138008924809