Run Time Assurance and Human AI Fluency in Crewed Autonomous Intelligence, Surveillance, and Reconnaissance

Richard Agbeyibor, Vedant Ruia, Carmen Jimenez Cortes, Jack Kolb
Adan Vela, Samuel Coogan, Karen Feigh

Richard Agbeyibor
PhD Student
Georgia Institute of Technology

AIAA Aviation Forum
1 August 2024

Georgia Tech

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

Abstract

The maturation of autonomy and electric Vertical Take-Off and Landing aircraft could soon make it possible to execute military Intelligence, Surveillance, Reconnaissance missions aboard crewed autonomous aerial vehicles.

This research experimentally investigates factors that may influence the quality of interaction between a non-pilot human operator and the AI pilot aboard such an aircraft. In a flight simulator study with twenty-seven participants, various levels of workload and AI capabilities are investigated including run time assurance. Control Barrier Functions are used to enable pro-active collision avoidance behaviors by an AI agent controlling the aircraft trajectory.

Team fluency and mission effectiveness outcomes show that trust, situation awareness, workload, perceived performance and user interface design are statistically significant factors for the quality of human AI interaction in this context.

Georgia Tech

Human-AI Collaboration in Autonomous Aerial Vehicles

Office of Naval Research Science of Autonomy
N00014-21-1-2759

PIs: Dr. Karen Feigh & Dr. Sam Coogan – GT
Dr. Adan Vela – UCF

Students: Maj Richard Agbeyibor, Mr. Vedant Ruia, Mr. Jack Kolb,
Ms. Sonya Doda, Ms. Carmen Jimenez Cortes

Research Goals and Objectives

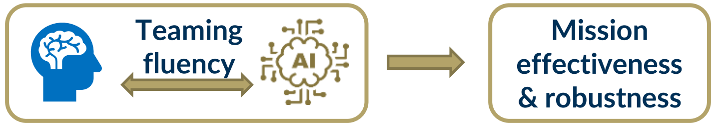
Autonomous aircraft

Onboard personnel with minimal AI training

Research Goal and Objectives:

- To understand elements of team fluency that are needed for an AI pilot to seek and receive assistance from on-board personnel with no direct training in either piloting or AI programming and for this personnel to team with an AI Pilot.

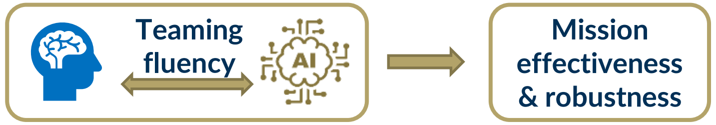
- Enable appropriate Human-AI collaboration needed to deal with off-nominal events by
  - (1) characterizing challenges to fluency
  - (2) quantifying the impact of fluency on mission effectiveness,
  - (3) exploring and validating mitigation strategies to improve fluency



Agenda

- Teaming Fluency
- Intelligence, Surveillance, and Reconnaissance context
- Emerging Opportunities in Advanced Air Mobility
- Study Scenario for Crewed Autonomous ISR
- Crewed Autonomous ISR Flight Simulator
- Study Protocol
- Results

Teaming Fluency



- Simply, Fluency is the quality of interaction in a team
- It has been described as the "elusive yet palpable characteristic that exists when two agents collaborate at a high degree of coordination and adaptability, particularly when they are habituated to the work of one another" (Hoffman and Brazeal, 2007)
- Literature on Fluency in Human-AI Collaboration focuses on
  - Turn-by-turn manufacturing context
  - Divisible, Maximizing and Additive tasks
- Literature on Human-AI Collaboration in autonomous vehicles focuses on
  - Autonomous cars teamed with licensed drivers
  - Autonomous aircraft teamed with professional pilots
- Little research on
  - Complex mission tasks that are highly interdependent
  - Non vehicle operators with critical mission responsibilities in autonomous vehicles
- We operationalize Fluency in our context through subjective and objective measures:
  - Mission effectiveness, User Interface actions
  - Situation Awareness, Workload
  - Trust, Perceived Performance

Intelligence, Surveillance, and Reconnaissance (ISR)

Simply, ISR is the task of **persistent monitoring of a target or area.**
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

Uncrewed ISR

ISR is a costly mission in terms of specialized aircraft, crew manning, and training requirements

Georgia Tech

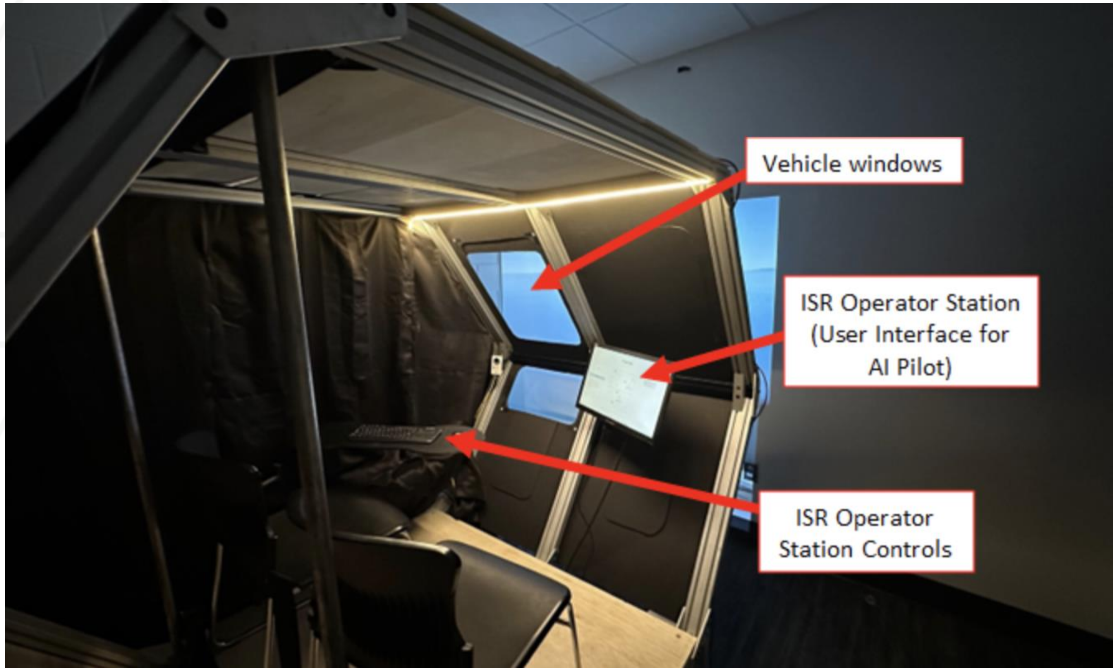
Emerging Opportunities in Advanced Air Mobility

Electric Vertical Take-Off and Landing (eVTOL) aircraft

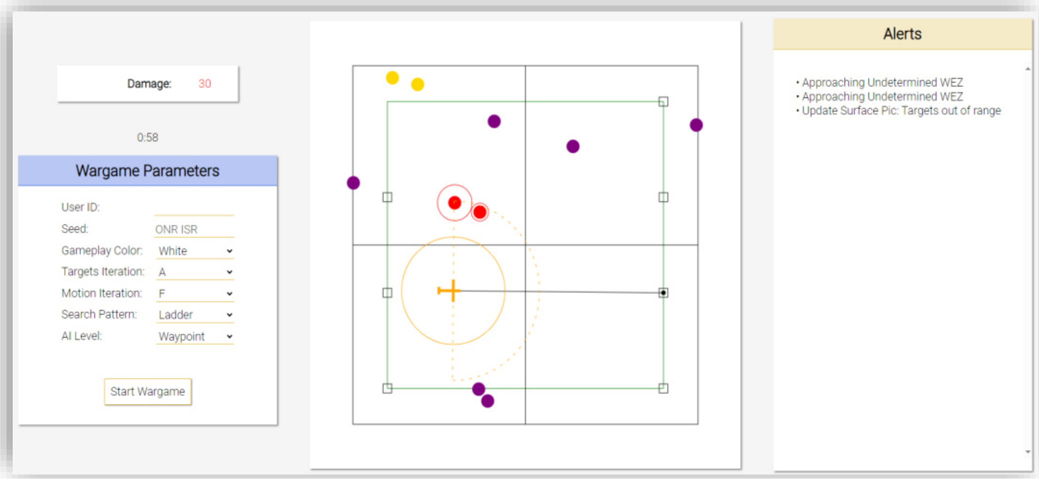
*Image used with permission from Volocopter GmbH

Aircraft Autonomy

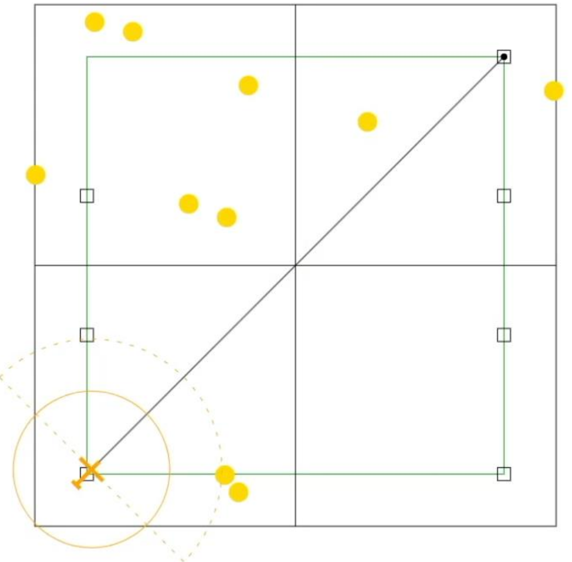
Scenario Design and Experimental Apparatus



Crewed Autonomous ISR Flight Simulator



ISR Wargame Operator Control Station



- An intel analyst / operator on-board is tasked to **collaborate** with an autonomous aircraft's **AI Pilot** to **identify and classify ships** in an assigned Surveillance Area while **avoiding damage** from enemy ship weapons
- 1 human + 1 AI dyad

User Study Protocol

Participant Briefing - Scenario Description

- Loss of RPA satellite control capabilities prompts military to put new recruit intelligence analysts aboard autonomous aircraft to conduct maritime patrol
- The mission is to classify all ships in an assigned surveillance area:
    - Minimize time
    - Minimize overflight of armed enemy ships

Participants

- Recruited from university population and local community
- 27 participants
    - 19 male; 6 female; 2 non-binary
    - 21 w/ no AI experience; 6 with
        - 1 self-taught
        - 1 undergraduate coursework
        - 4 graduate coursework
    - 22 w/ no flight experience; 5 with
        - 4 with less than 10 hours
        - 1 licensed private pilot

Experimental Design

- Participant collaborates with AI pilots with 4 different AI behavior levels on missions with 2 task load levels – for a total of 8 scenarios
- Full-factorial, counterbalanced, within-subjects experimental design
- Evaluate team fluency through
    - Mission effectiveness
    - Situation Awareness
    - Workload
    - Trust

Georgia Tech

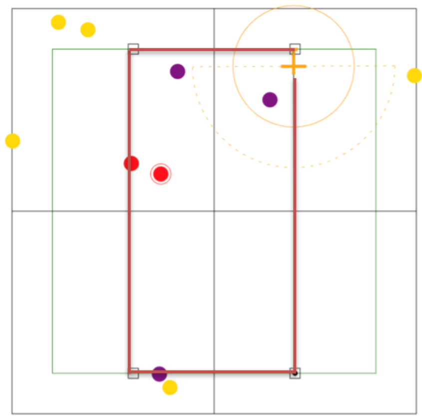
AI Behaviors and Task Load

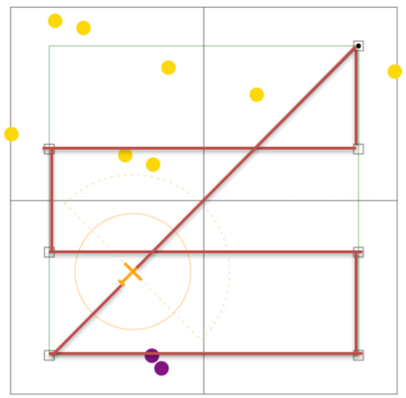
AI Behaviors

- Waypoint
  - Automated search pattern navigation
  - AI always accepts user waypoints
- Collaborative
  - AI accepts or denies user waypoints
- Collision Avoidance (Run Time Assurance)
  - AI suggests obstacle avoidance waypoints to user
  - Without user input, AI proactively avoids red obstacles
  - AI blends in user waypoints with collision avoidance
- Search Optimization
  - User can request AI suggestions for search pattern optimizations
  - User accepts or vetoes AI suggestions

Task Load

- Low
  - 10 targets
  - Square Search Pattern
- High
  - 20 targets
  - Ladder Search Pattern





Run Time Assurance

Control Barrier Functions (CBFs)

- CBFs used as RTA mechanism in Collision Avoidance autonomy behavior
- Enforces forward invariance of a constraint set so that no trajectory violates the safety specifications for the system
- Minimizes risk of aircraft damage and enforces safe trajectories

Implementation in ISR Wargame

- CBFs implemented a second order unicycle model to simulate the aircraft
- CBF-enabled controller took as inputs target ship locations and the size of their Weapon Employment Zones
- Built barriers around each one of them and modified aircraft trajectory to avoid collision with these barriers

Georgia Tech

Results – RQ1: Task Complexity vs Fluency

How do changes in task complexity affect situation awareness, workload and mission effectiveness?

- Hypothesis 1: An increase in task complexity will decrease situation awareness, increase workload, and decrease mission effectiveness

- (H1) Workload, situation awareness, mission performance change as hypothesized, validating experimental design

TLX Subscale	Δ Low to High	p-value
Mental Demand	9.9	$p < 0.0001$***
Physical Demand	1.7	$p = 0.018$*
Temporal Demand	7.0	$p = 0.00006$***
Effort	11.5	$p < 0.0001$***
Frustration	13.1	$p < 0.0001$***
Performance	-16.4	$p < 0.0001$***

Table 2 Variance in Workload per Task Load

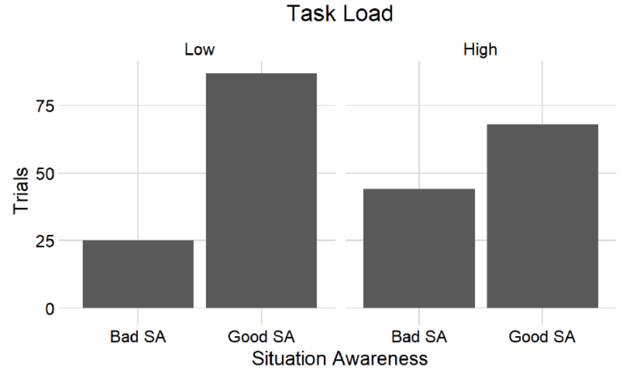
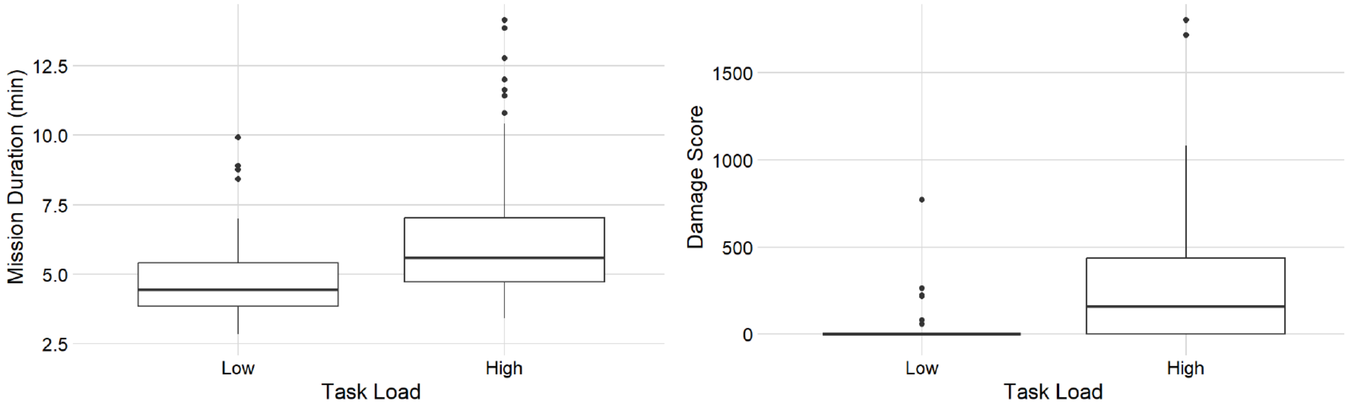


Fig. 2 Situation Awareness vs Task Load



(a) Mission Duration vs Task Load **(b) Aircraft Damage vs Task Load**

Fig. 5 Mission Effectiveness vs Task Load

Results – RQ2: AI Behavior vs Fluency

How do various autonomy behaviors such as CBF-enabled run time assurance, affect fluency components – situation awareness, perceived performance, interaction and workload?

- Hypothesis 2a: Levels of autonomy that increase decision support, such as CBF-enabled run time assurance, will decrease workload and perceived performance
- Hypothesis 2b: Levels of autonomy that share decision authority without transparency will increase workload, decrease situation awareness, and perceived performance

- (H2a) Participants gave higher ratings to CBF-enabled run time assurance AI behavior
- (H2b) Lack of transparency and explainability caused Frustration when AI changed user trajectory inputs
- Under high task load, the search optimization behavior which requires the operator to evaluate AI suggestions, had the highest SA

Fig. 6 Situation Awareness vs AI Behavior

TLX Subscale	p-value
Mental Demand	$p > 0.05$
Physical Demand	$p > 0.05$
Temporal Demand	$p > 0.05$
Effort	$p > 0.05$
Frustration	$p = 0.0005$**
Performance	$p = 0.0019$*

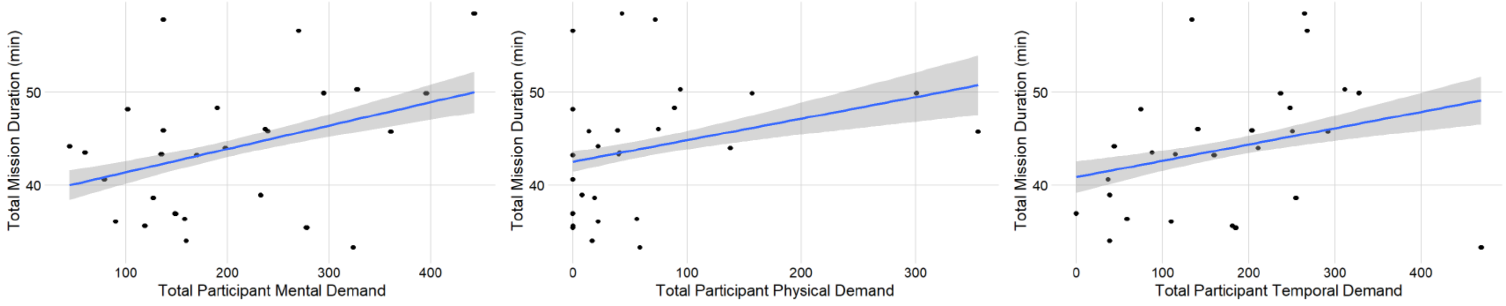
Fig. 7 Variance in Workload per AI Behavior

Results – RQ3: Fluency vs Mission Effectiveness

How do the fluency components – trust, situation awareness, perceived performance, interaction and workload – affect mission effectiveness?

- Hypothesis 3: A decrease in fluency , indicated by an increase in trust, situation awareness, and perceived performance, along with a decrease in interaction and workload, will decrease mission effectiveness

- (H3) General trends of positive correlation between fluency and mission effectiveness on all scales



(a) Mental Demand (b) Physical Demand (c) Temporal Demand

Fig. 11 Total Mission Duration vs Workload

Conclusions

- With minimal training, participants learned to collaborate with various AI pilots to accomplish crewed autonomous ISR
- Higher number of targets and more complex search patterns increased workload
- Higher workload was correlated with decreased situation awareness, trust, performance
- AI interaction mechanism that required human review and approval of AI decisions yielded higher situation awareness
- Lack of transparency and explainability of AI decisions led to high frustration
- Run time assurance through CBF-enabled collision avoidance, led to lower aircraft damage but higher mission duration
- Team fluency was positively correlated with mission effectiveness on all scales

Discussion

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