Human AI Collaboration

Richard Agbeyibor
Robotics PhD Student
Georgia Institute of Technology

Georgia Tech

Agenda

- Teaming
- Autonomous RPA
- Autonomous Wingman
- Crewed Autonomy
- Medical Evacuation
- Intelligence Surveillance Reconnaissance

Cognitive Engineering Center

The CEC continues to ignore the boundaries of traditional disciplines in our search for meaningful, implementable solutions. Aerospace engineers, computer scientists, roboticists, industrial engineers, and education researchers work together to build a safer and more effective human-machine world.

In our complex world of humans and machines, Cognitive Engineering Center researchers in the Feigh Research Group are building the foundations, training, and technologies for safe and effective work.

Dr. Karen Feigh and Aerospace Engineering student Vedant Ruia use a voice-activated Google assistant to help with pilot checklists and troubleshooting in emergency scenarios (e.g., smoke in cabin, pilot incapacitation)(2023).

Grad students Robert Walters and Vinodhini Comandur are among several researchers who have already begun to use the rotorcraft simulation lab to investigate new techniques for improving rotorcraft safety. Credit: Daniel Guggenheim School of Aerospace Engineering (2018)

Human Autonomy Interaction

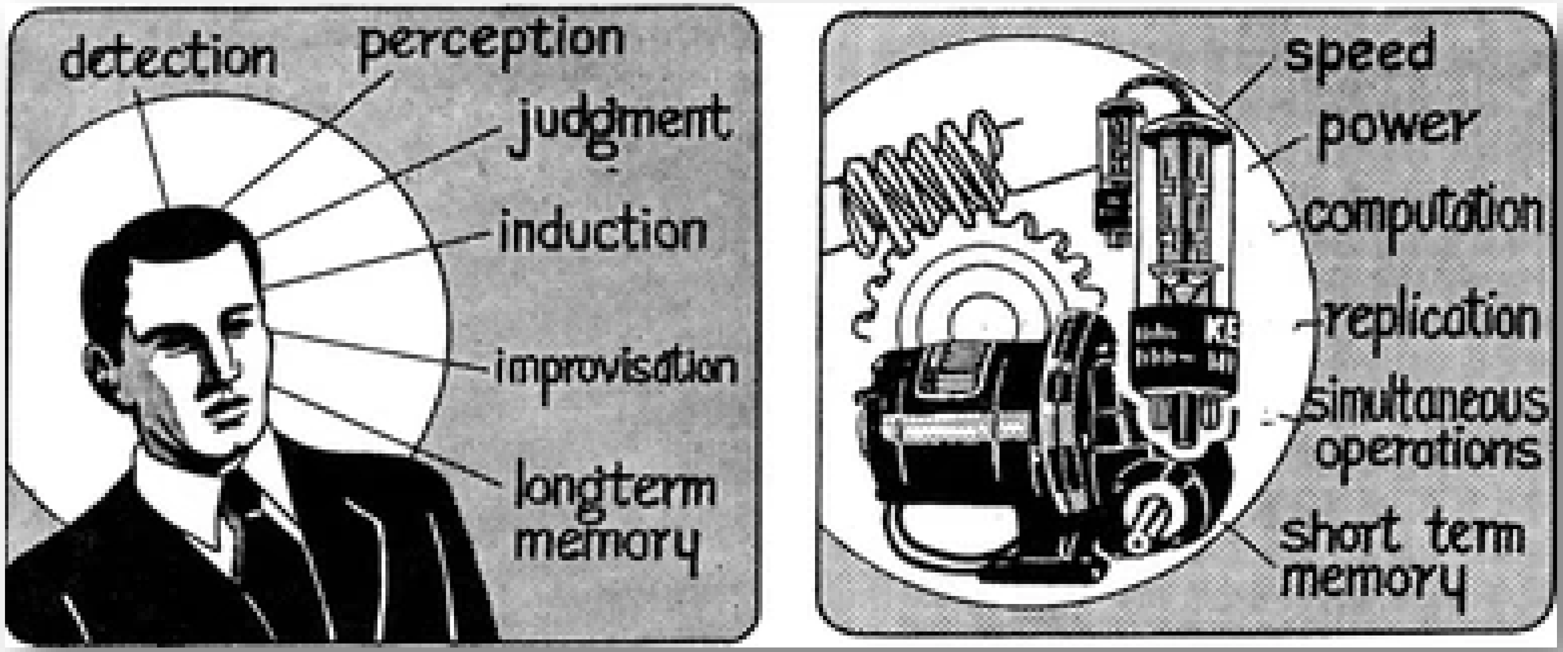
Interaction, *the process of working together to accomplish a goal*

Understanding, designing, and evaluating autonomy, robotic, and machine systems for use by or with humans, in various domains.

Space	Command and Control	Military Aviation	General Aviation

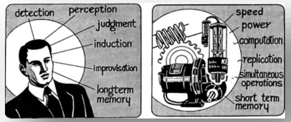
Georgia Tech

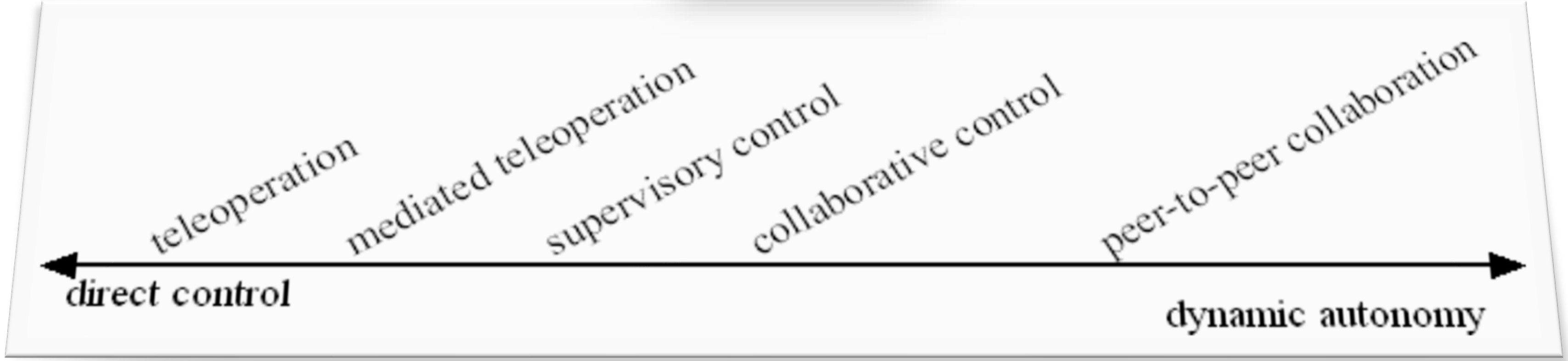
Human Autonomy Teaming

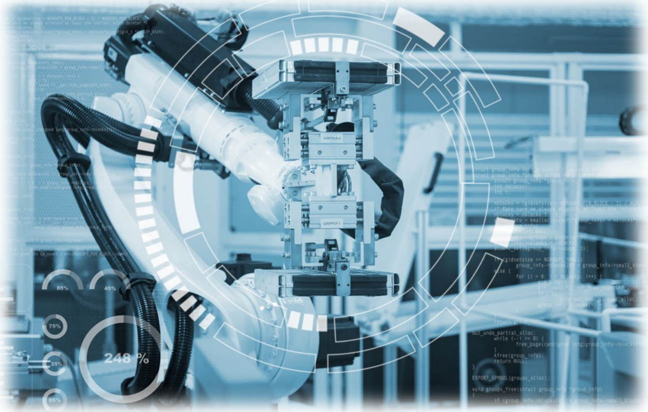


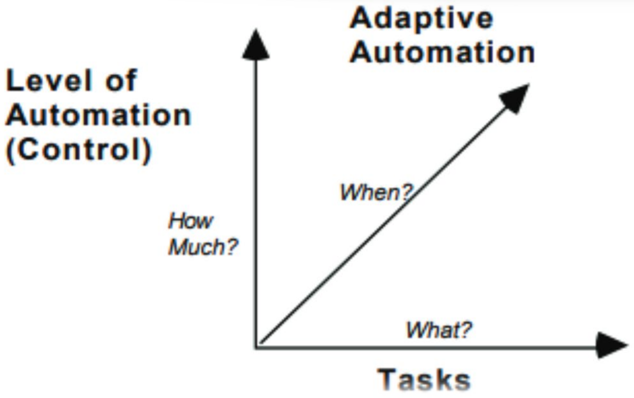
Fitts' List, Men Are Better At, Machine Are Better At, 1951

Human Autonomy Teaming

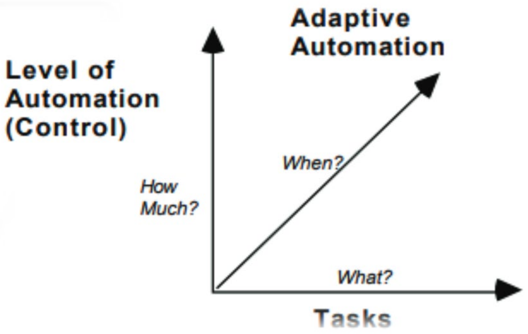


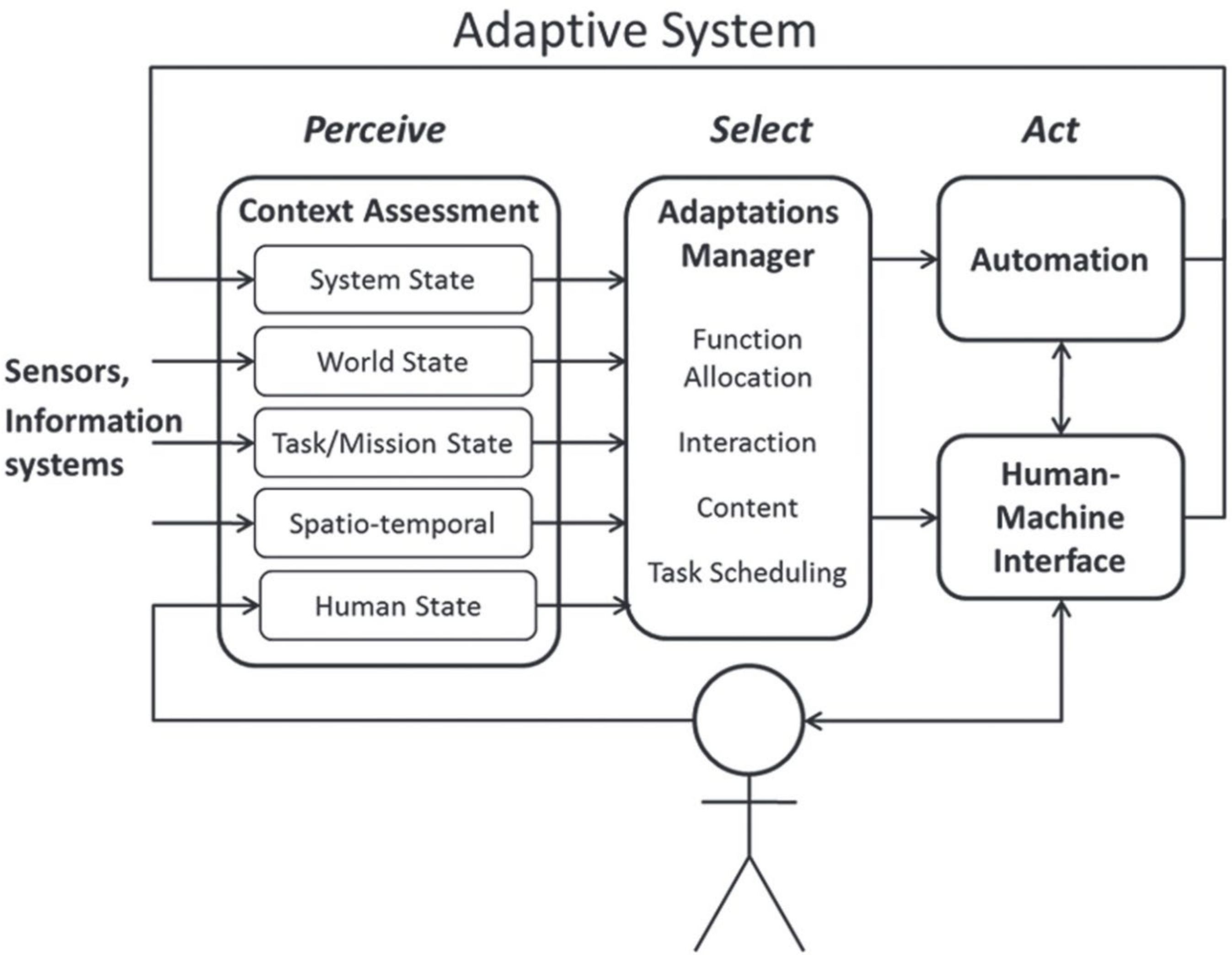






Human Autonomy Teaming





Autonomy in the cockpit: Remote Piloted Aircraft

An aircrew from the California Air National Guard's 163rd Attack Wing flies an MQ-9 Reaper remotely piloted aircraft during a mission to support state agencies fighting the Mendocino Complex Fire in Northern California, Aug. 4, 2018. The aircrew conducted fire perimeter scans and spot checks on the blaze, which encompasses the Ranch and River fires. California Air National Guard photo by Senior Airman Crystal Housman https://www.defense.gov/News/News-Stories/Article/Article/1595204/

Georgia Tech

Autonomy in the cockpit: Manned Unmanned Teaming

An AH-1Z Viper (top) with Marine Operational and Test Evaluation Squadron 1 (VMX-1), and an MQ-8C Fire Scout unmanned helicopter assigned to Helicopter Sea Combat Squadron 23 (HSC-23), conduct Strike Coordination and Reconnaissance Training near El Centro, California, March 10, 2022. The purpose of this exercise was to provide familiarization and concept development of manned-unmanned teaming. (U.S. Marine Corps photo by Lance Cpl. Jade Venegas)
https://www.navalnews.com/naval-news/2022/03/usmc-and-us-navy-demonstrate-manned-unmanned-teaming/

Autonomy in the cockpit: Crewed Autonomous Aircraft

Future Scenario:	Aircraft control & piloting will be controlled by embedded AI pilot
	AI pilot will competently handle nominal and basic off-nominal conditions
	Onboard personnel will not be trained to pilot or manually control aircraft
	However, personnel will have unique understanding of mission & situational factors to guide/override AI

What can we reasonably expect?

Human-AI Collaboration in Autonomous Aerial Vehicles

N00014-21-1-2759

PIs: Dr. Karen Feigh & Dr. Sam Coogan – GT; Dr. Adan Vela – UCF
Maj Richard Agbeyibor, Ms. Sanya Doda, Mr. Jack Kolb,
Ms. Carmen Jimenez Cortes

Cognitive Engineering Center

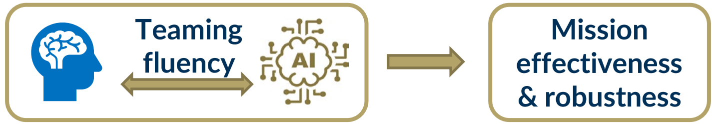
Research Goals and Objectives

Autonomous aircraft

Onboard personnel with minimal AI training

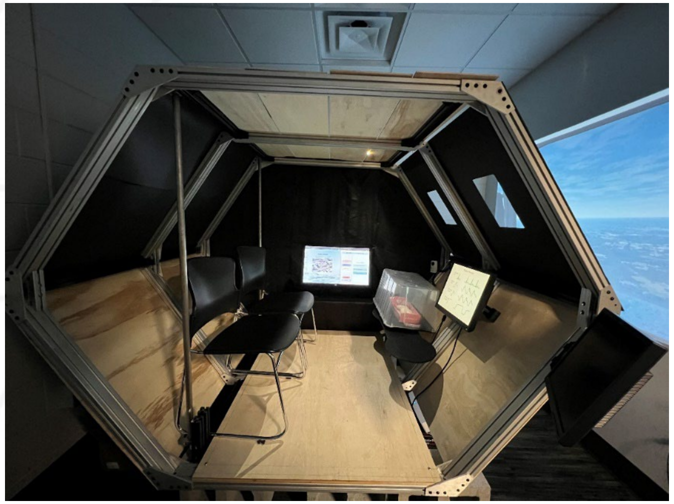
Research Goal and Objectives:

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

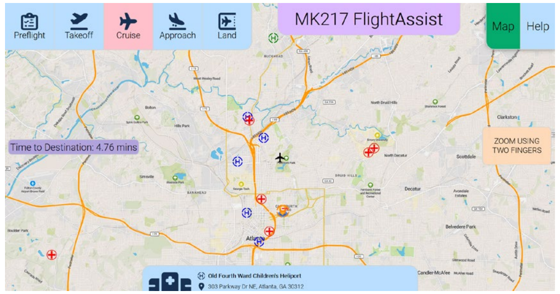


Georgia Tech

Autonomous Medical Evacuation



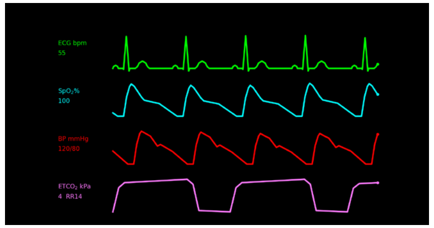
Medevac Simulator Cabin

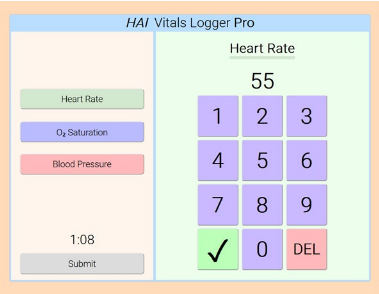


Touchscreen UI
for Human-AI Pilot Interaction

Scenario Description

- Premature infant born at 29 weeks requires incubation. Medical evacuation for heart surgery at Scottish Rite at Rush Hour.
- Weather is overcast reducing visibility.
- Off-nominal events incorporated:
  - Unmarked landing zone obstacles
  - Change in weather
  - Change in patient status





UI for Patient Vitals Monitoring and logging

User Study

Designed four Medevac Scenarios:
- Scenario 1: Baseline
- Scenario 2: Landing Zone Obstacle
- Scenario 3: Weather Emergency
- **Scenario 4: Patient Emergency**

Recruited participants:
- With emergency medical technician, paramedic training or equivalent (anytime in the past 10 years).
- Instrumented with heart rate, respiratory rate, eye tracker, hand tracker and other physiological sensors

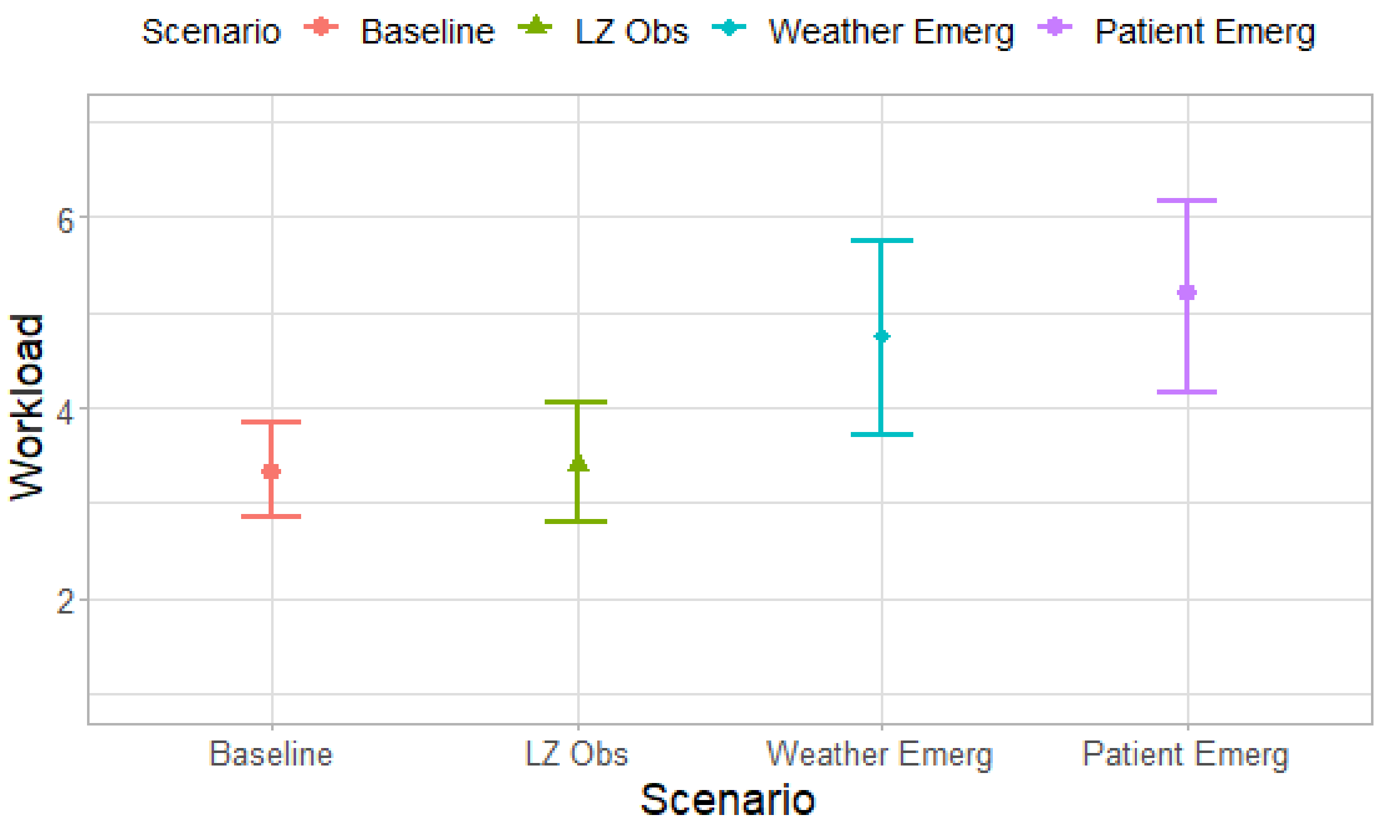
Snippet from Scenario 4: Patient Emergency

Preliminary Results

Recruited 10 participants (9 EMTs, 1 MD) with no piloting or AI programming experience

They successfully accomplished simulated Medevac missions with an autonomous aircraft

The workload seemed reasonable throughout – but higher with complex scenarios/interaction needs



Scenarios induced differential workload.

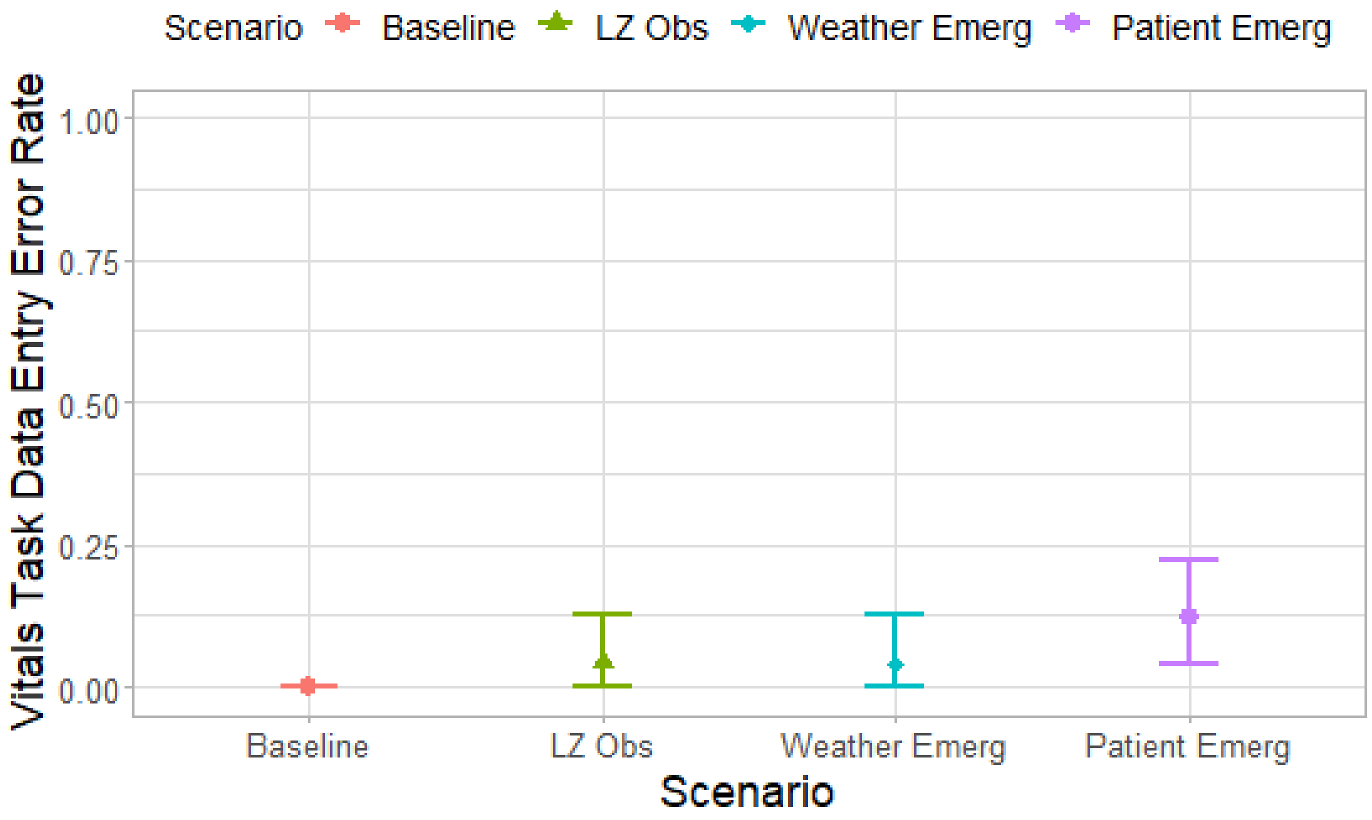
Preliminary Results

Errors were made, but not in a way that indicated severe incapacity to function

AI-Pilot influence detected, but not easily classified as over-reliance

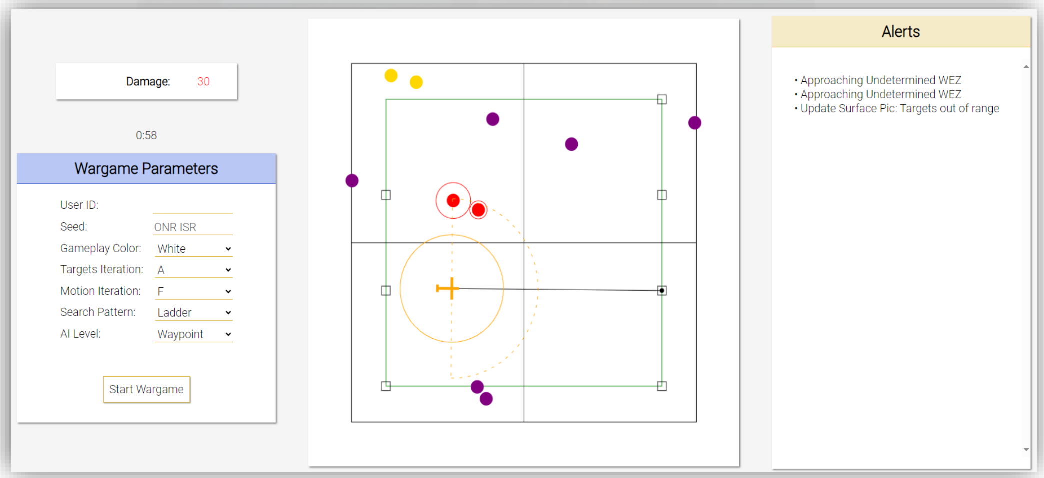
In open-ended questions, subjects requested:

- Better transparency into the AI's decision-making process
- Speech interface with the AI so that they could keep their hands free for medical tasks
- Training of the AI on the medical context and its critical phases



We start to see increases in error rates with scenario complexity and workload.

Manned Intelligence Surveillance Reconnaissance (ISR) Aboard Autonomous Aerial Vehicles



ISR Wargame Operator Control Station

Scenario Description

- Manpower shortage and maturity of autonomous eVTOLs prompts military to put non-pilot intelligence analysts aboard autonomous aircraft to conduct maritime patrol
- Task is to classify all ships in a surveillance area
  - Minimize time
  - Minimize overflight of armed enemy ships
- Analyst collaborates with an AI pilots with different behaviors on missions with various task loads
- Evaluate team fluency through
  - Mission effectiveness
  - Situation Awareness
  - Workload
  - Trust

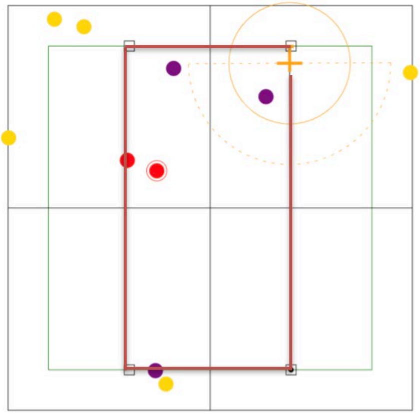
AI Behaviors and Task Load

AI Behavior

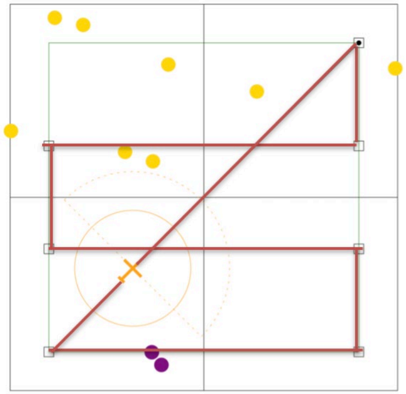
- Waypoint
  - Automated search pattern navigation
  - AI always accepts user waypoints
- Collaborative
  - AI accepts or denies user waypoints
- Collision Avoidance
  - AI suggests obstacle avoidance waypoints to user
  - Without user input, AI proactively avoids red obstacles
  - AI blends in user 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

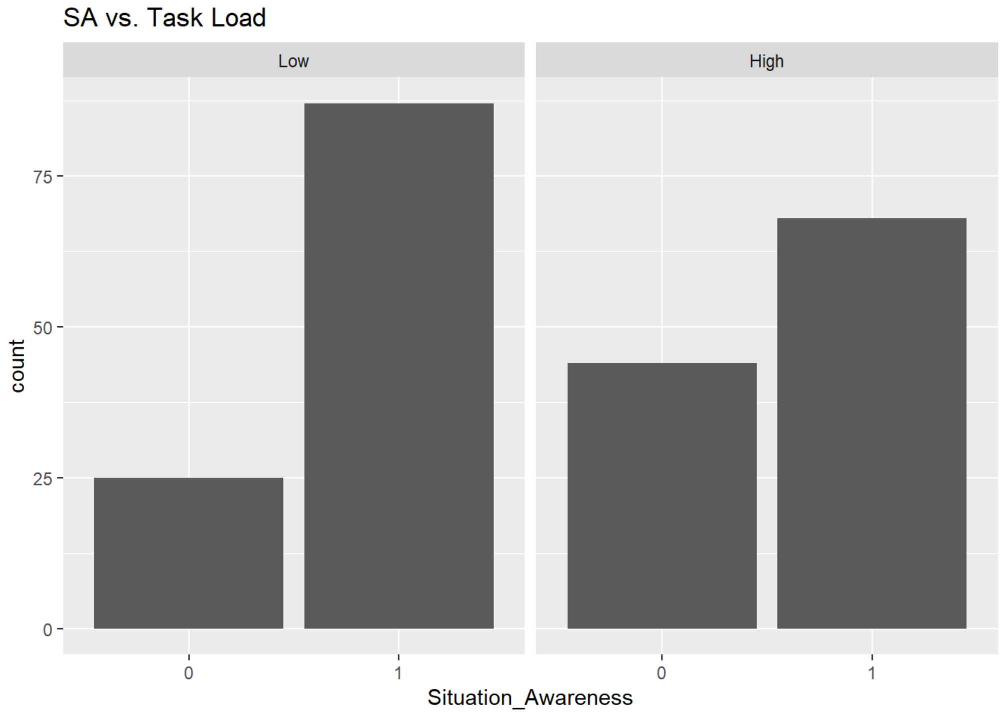


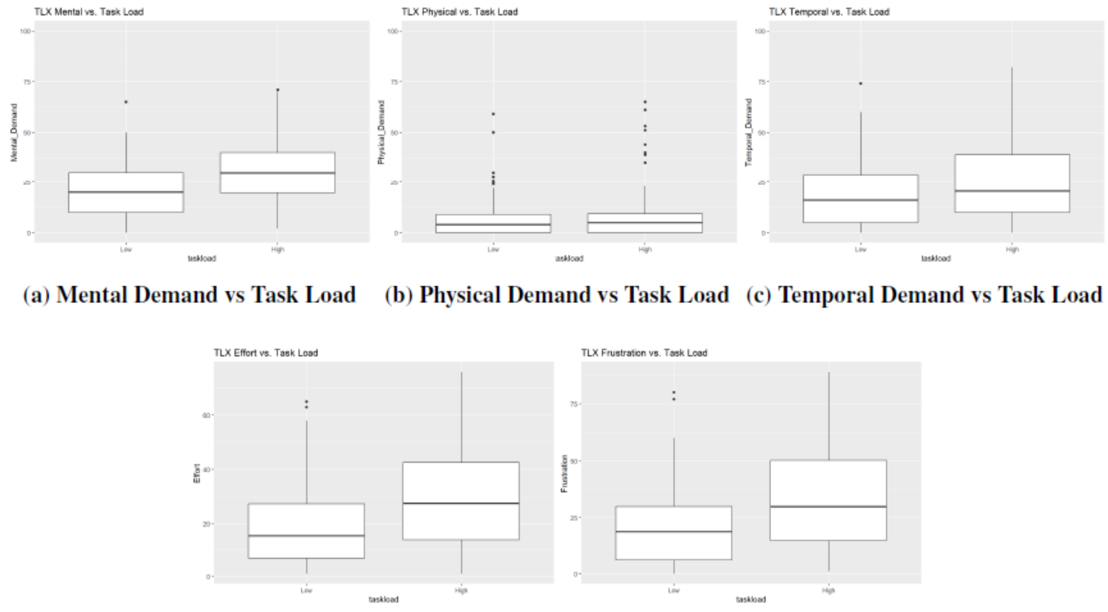
- High
  - 20 targets
  - Ladder Search Pattern



Preliminary Results

- RQ1:Task Complexity vs Fluency
  - Task load decreased situation awareness
  - NASA TLX showed increased workload with increased task Load





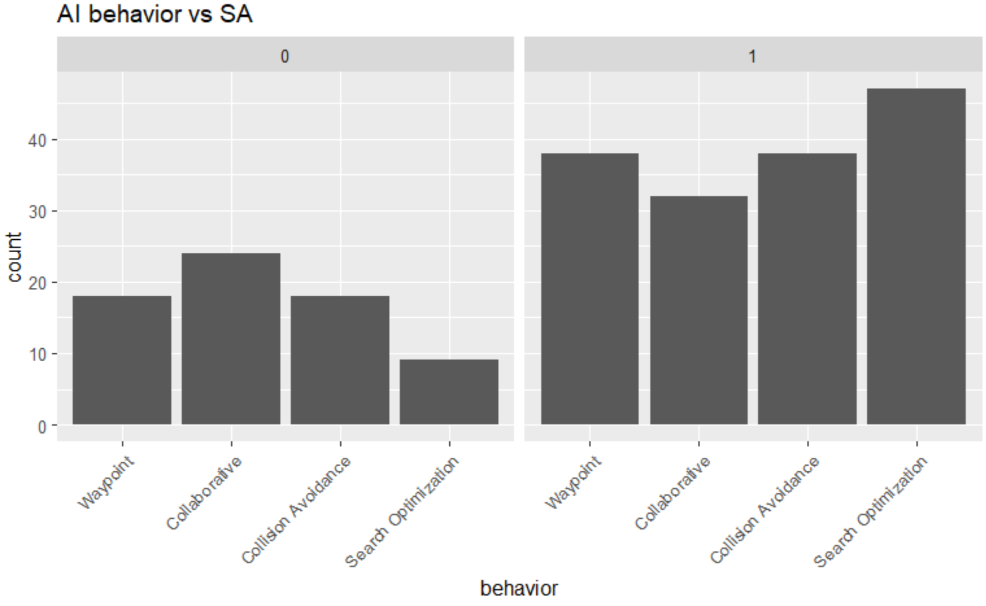
(a) Mental Demand vs Task Load (b) Physical Demand vs Task Load (c) Temporal Demand vs Task Load

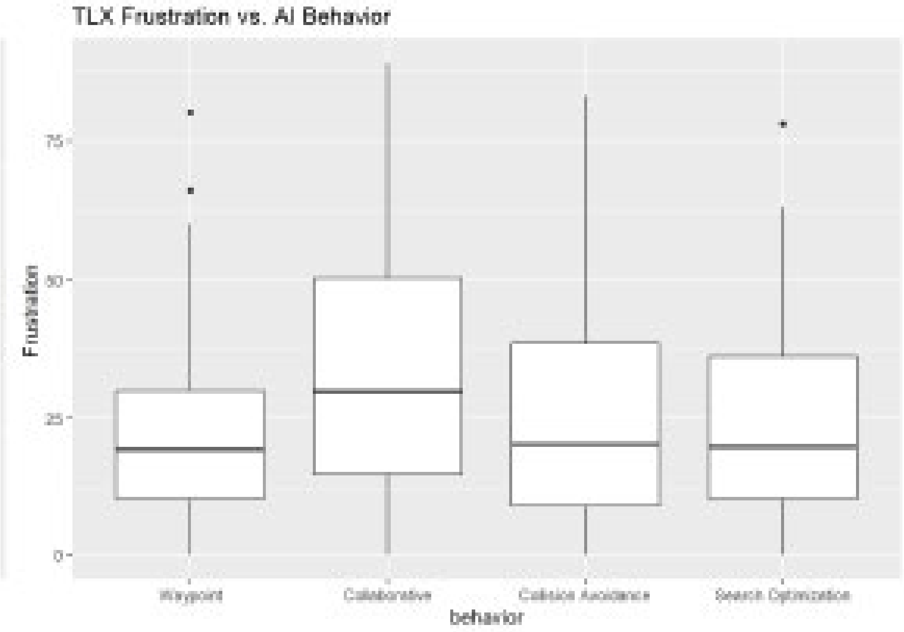
(d) Effort vs Task Load (e) Frustration vs Task Load

Georgia Tech

Preliminary Results

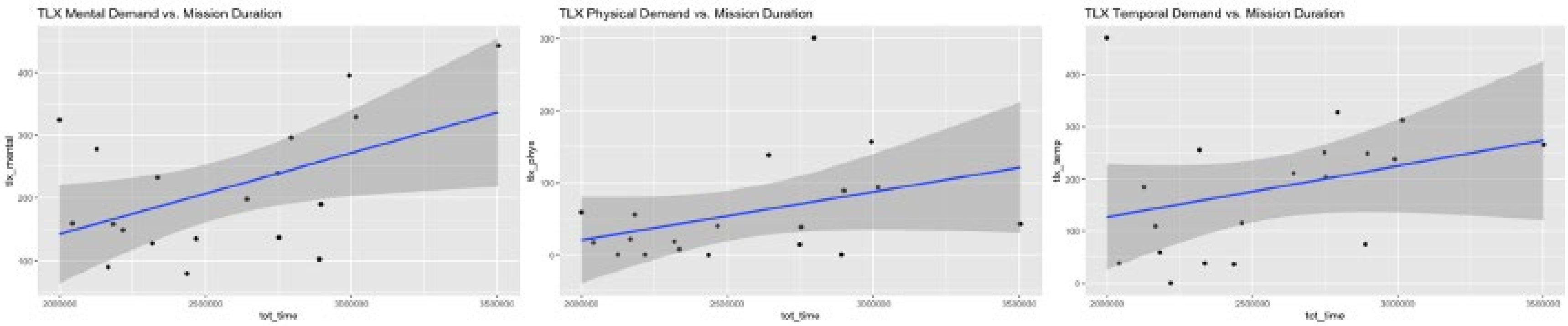
- RQ2: AI Behavior vs Fluency
  - Situation awareness was best when user was required to evaluate AI suggestions (Search Optimization AI)
  - Frustration increased when AI denied user veto (Collaborative AI)





Preliminary Results

- RQ3: Fluency vs Mission Effectiveness
  - High workload correlated with low mission effectiveness
  - No significant relationship between trust ratings and mission effectiveness
  - No significant relationship between physiological measures and mission effectiveness



- Open ended questions
  - Participants with AI experience tended to be very skeptical of AI, and expressed mistrust
  - Participants with no AI experience tended to attribute higher capabilities to AI, and expressed overtrust

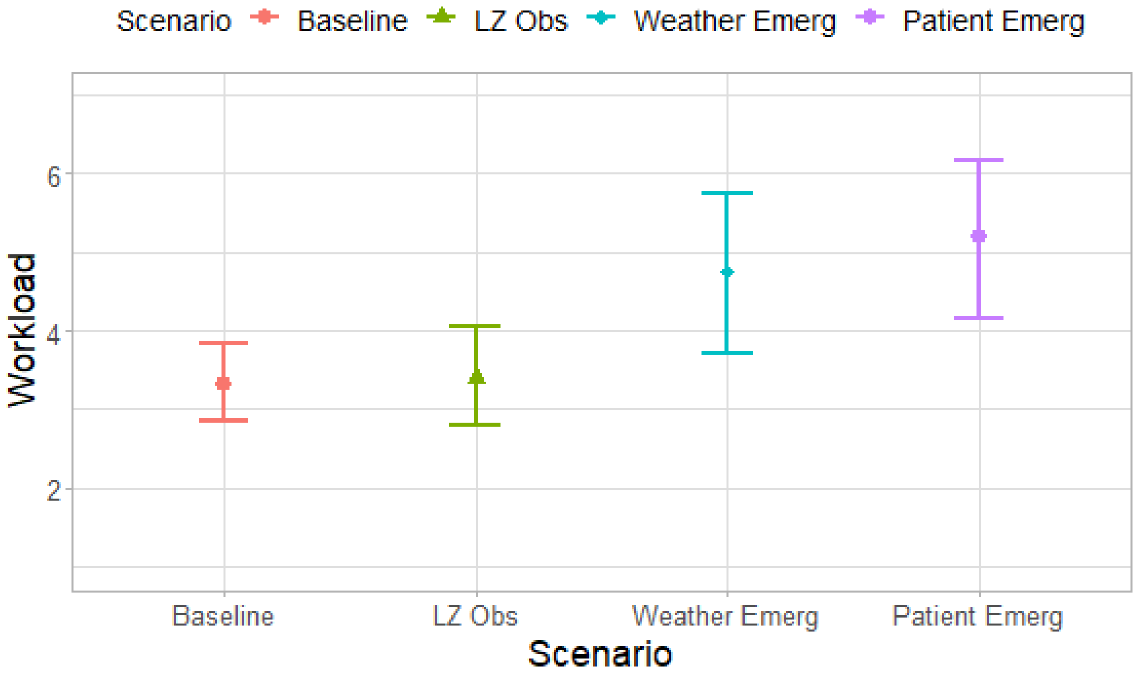
Georgia Tech

Questions?

BACK UP SLIDES

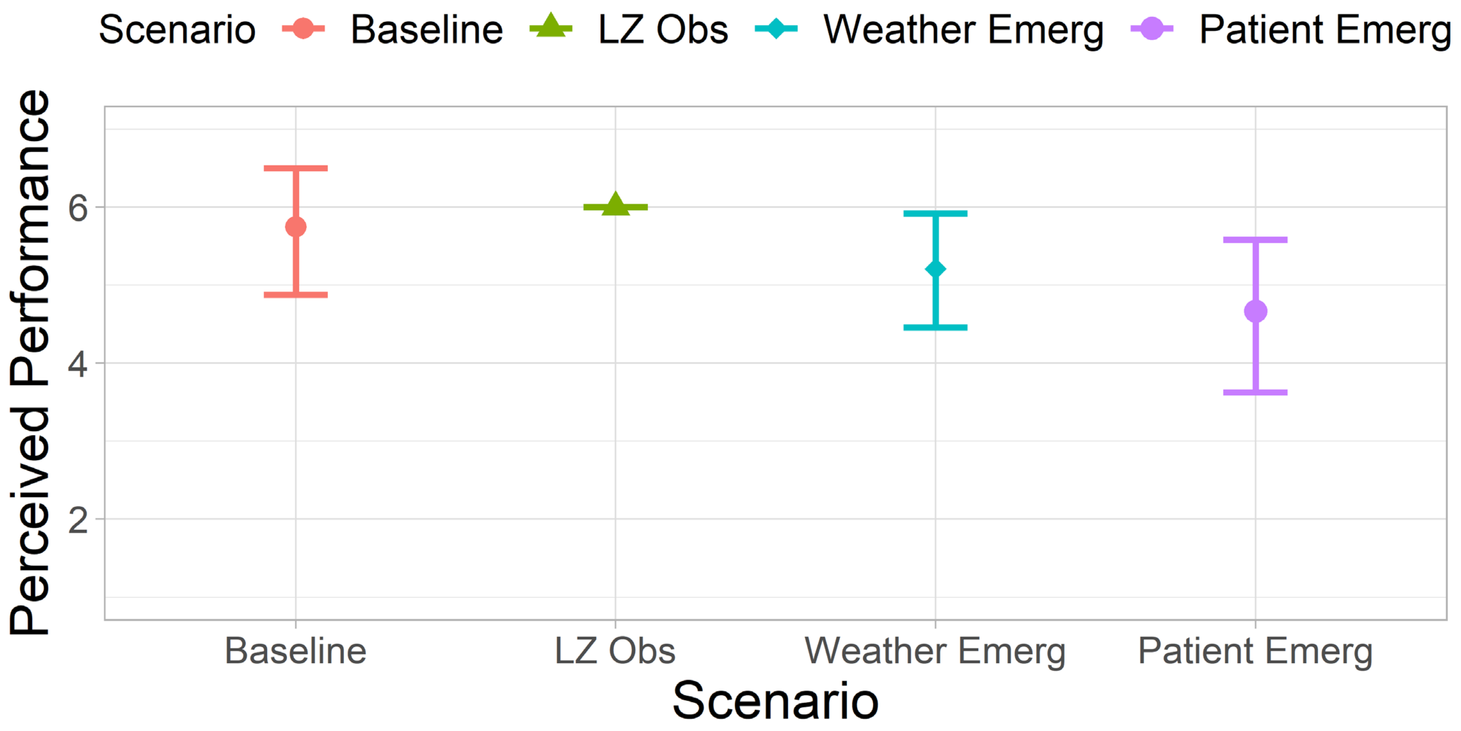
Subjective Measures

- Modified NASA TLX:
  - Workload
    - 7-point rating scale
    - Mean of unweighted subscales



Scenarios were able to induce differential workload.

Perceived Performance



Perceived performance decreased as scenario complexity / AI interaction increased.

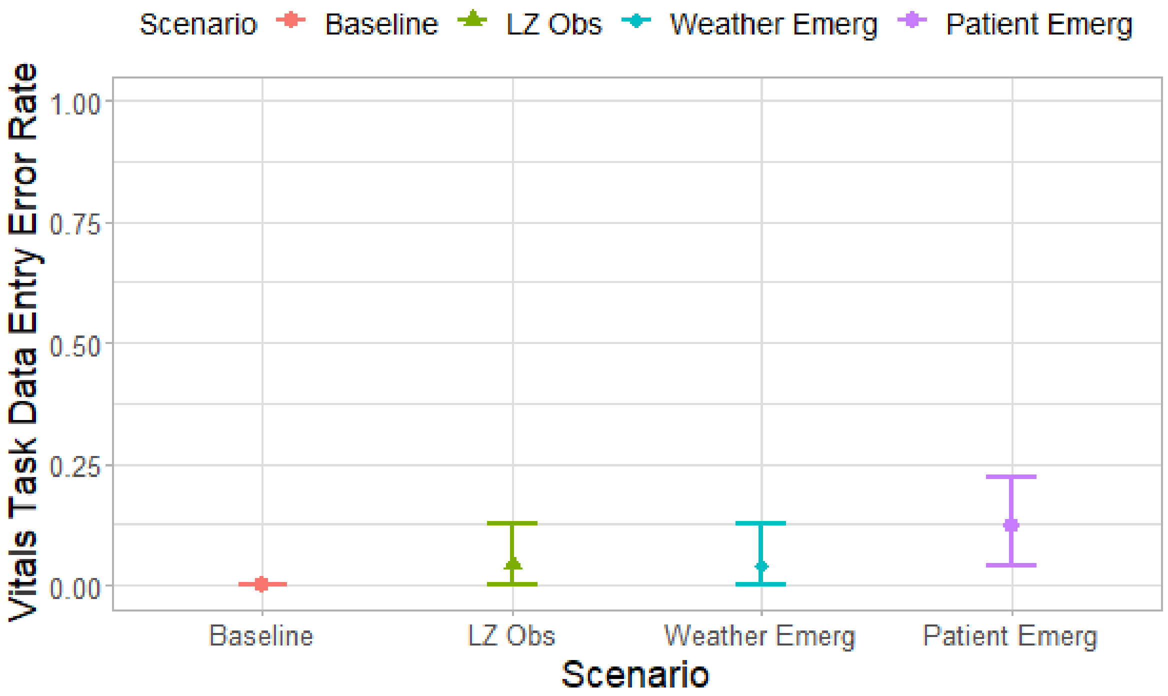
Medical Task Performance

Vitals Logging Task:

- Simulation of EMT enroute care through frequent vitals logging task
- Task requires 4 button presses & data entry for:
  - Patient Heart Rate
  - Patient Oxygen Saturation
  - Patient Blood Pressure

Task Performance Metrics

- Interval between prompt display and first button press
- Task duration
- Data Entry Error Rate



We start to see decreases in performance metrics with scenario complexity and human workload.

ISR Pilot Study

Preliminary Results

- RQ1:Task Complexity vs Fluency
  - Task load decreased situation awareness
  - NASA TLX showed increased workload with increased task Load

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	$p < 0.0001$***
Perceived Performance	-16.4	$p < 0.0001$***

Table 2 Variance in Workload per Task Load