Improving Human Situation Awareness in AI-Advised Decision Making

Divya Srivastava
Dr. Karen Feigh
November 19, 2022

Cognitive Engineering Center
Georgia Tech

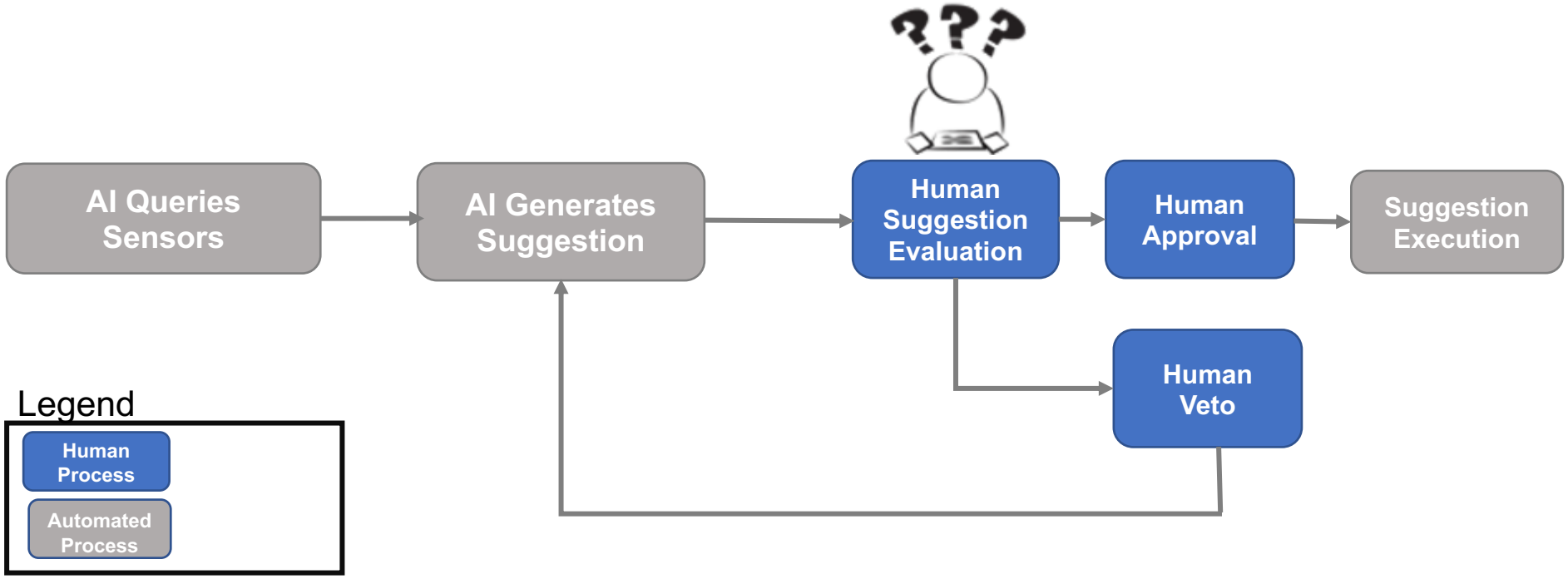
Human-AI Decision Making Teams

Aviation: Autonomous Mission Planners[1], HUDs, sensor suites

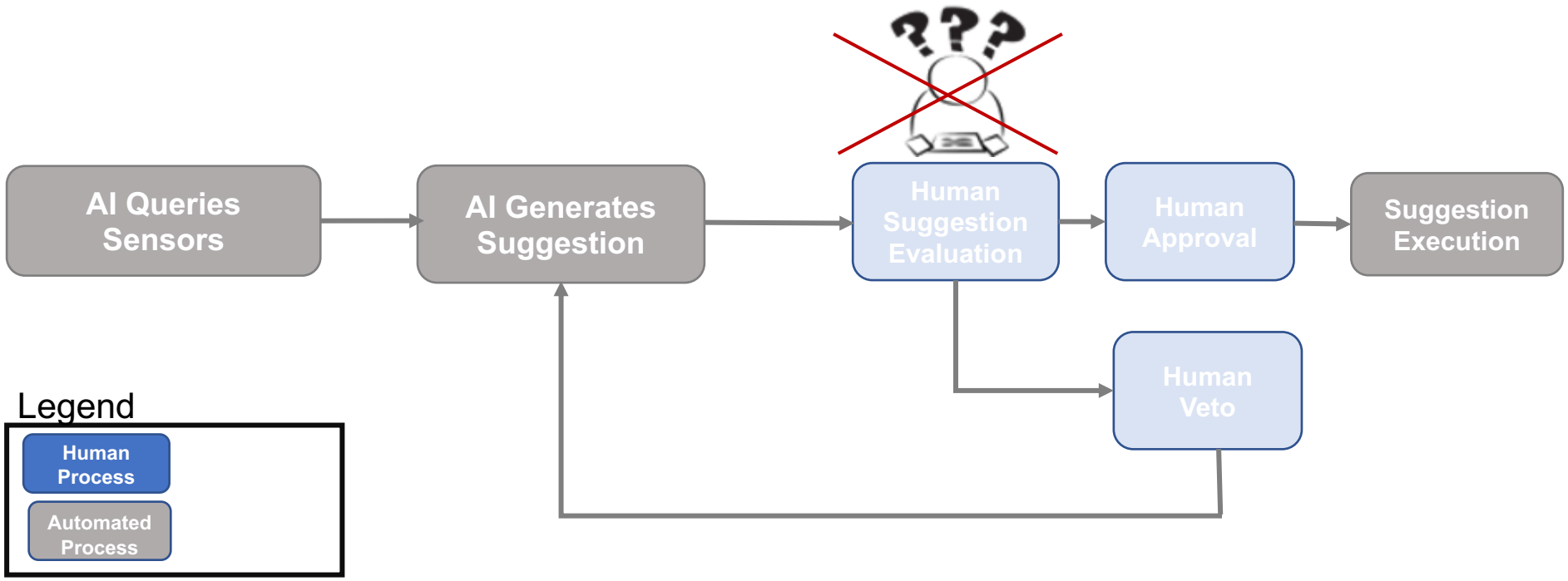
Medicine: diagnoses[3], active patient care

Finance: Loan eligibility[2], mortgage rate offer

AI-Advised Decision Making Process

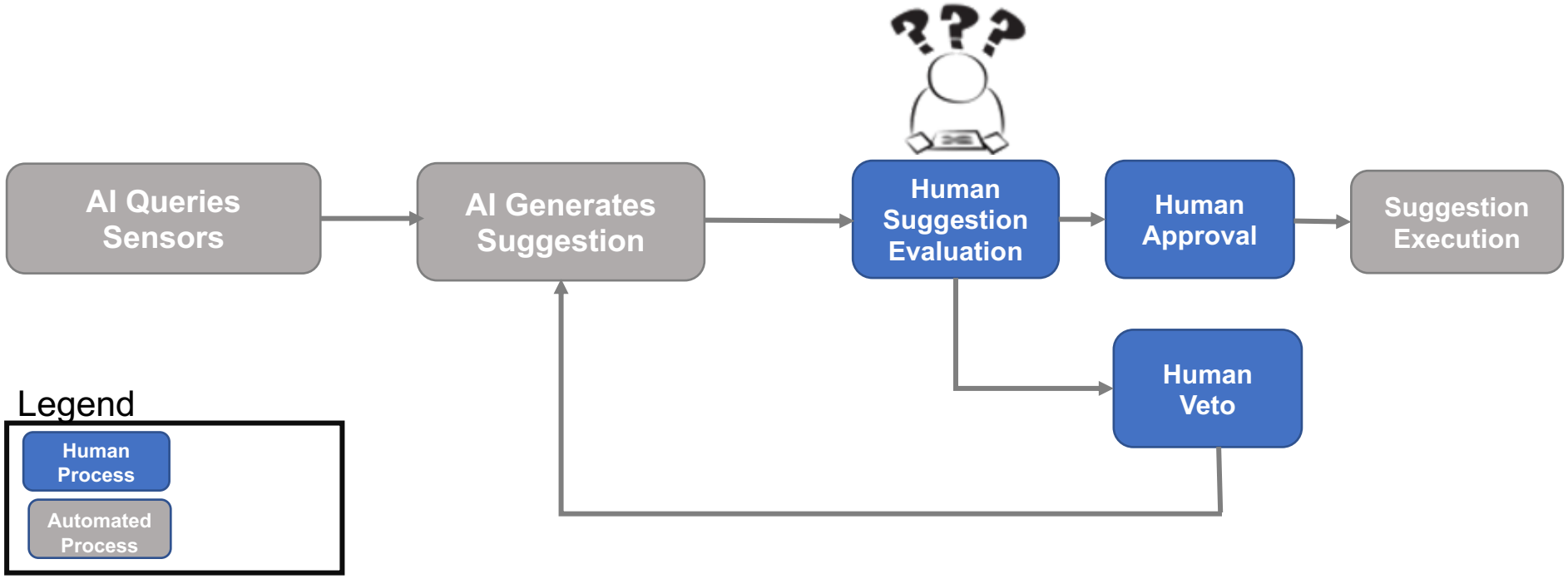


AI-Advised Decision Making Process in an **Emergency**



Humans have a tendency to over-rely on automation in emergencies.

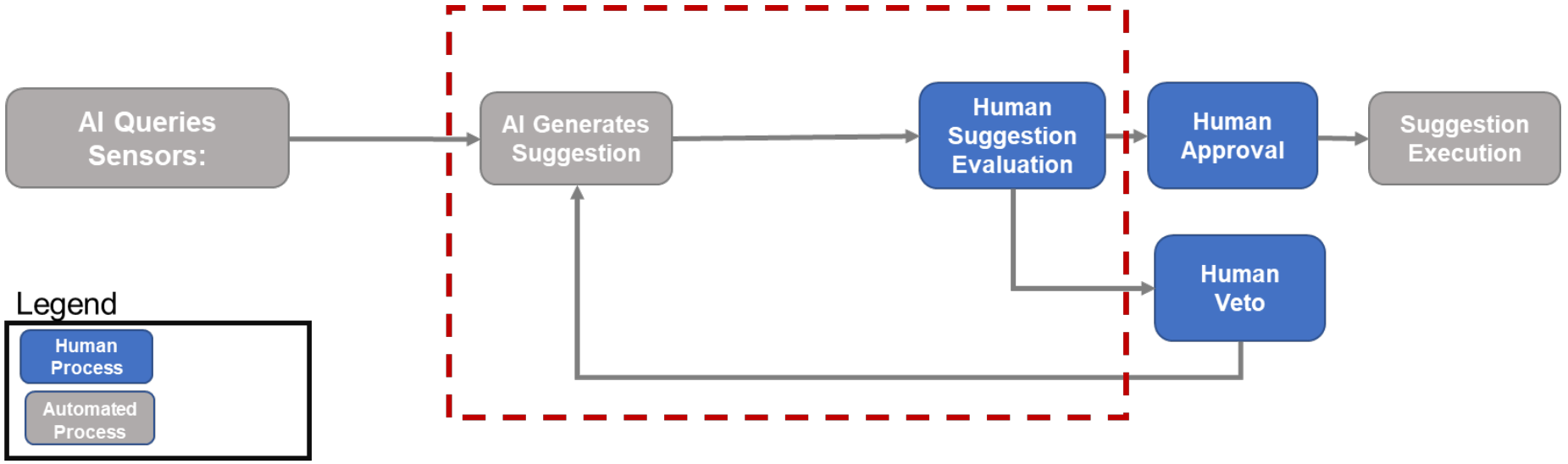
But going back to this framework, because it appears so often…



Can the human operator be successful in their safety & oversight role, or are we setting them up for failure?

Particularly in the face of black-box AI.

Current Approaches to Support the Human in this Framework



Suggestion Evaluation

- Increased interaction
- Visualization cues

Algorithmic Transparency

- Local Explanations [1, 2, 4]
- Global Explanations [2, 3, 4]
- Post-Hoc Feature Explanation [5]

[1] Lakkaraju et al, 2017
[2] Panigutti et al, 2022
[3] Koh and Liang, 2017
[4] Guidotti et al, 2018
[5] Slack et al, 2020

Limitations of Algorithmic Transparency/XAI

- Interpretability
  - no consensus on a "good" explanation
  - what helps one user may not help another
- As autonomy becomes more sophisticated, extracting a straightforward human-like explanation becomes difficult
- Too costly (time, labor, money, etc) to change existing systems

Current Study Objectives

If we cannot change the way AI is currently being deployed as a black box / and limited capacity to increase transparency in current AI systems, can we ***introduce transparency*** in the system in other ways such that it ***improves their collaborative performance*** with their AI partner?

Ideas and Inspiration #1 -- SMM Framework

Shared Mental Model: a shared perception of goals and actions through effective communication and an understanding of their fellow team members' goals and likely methods [1]

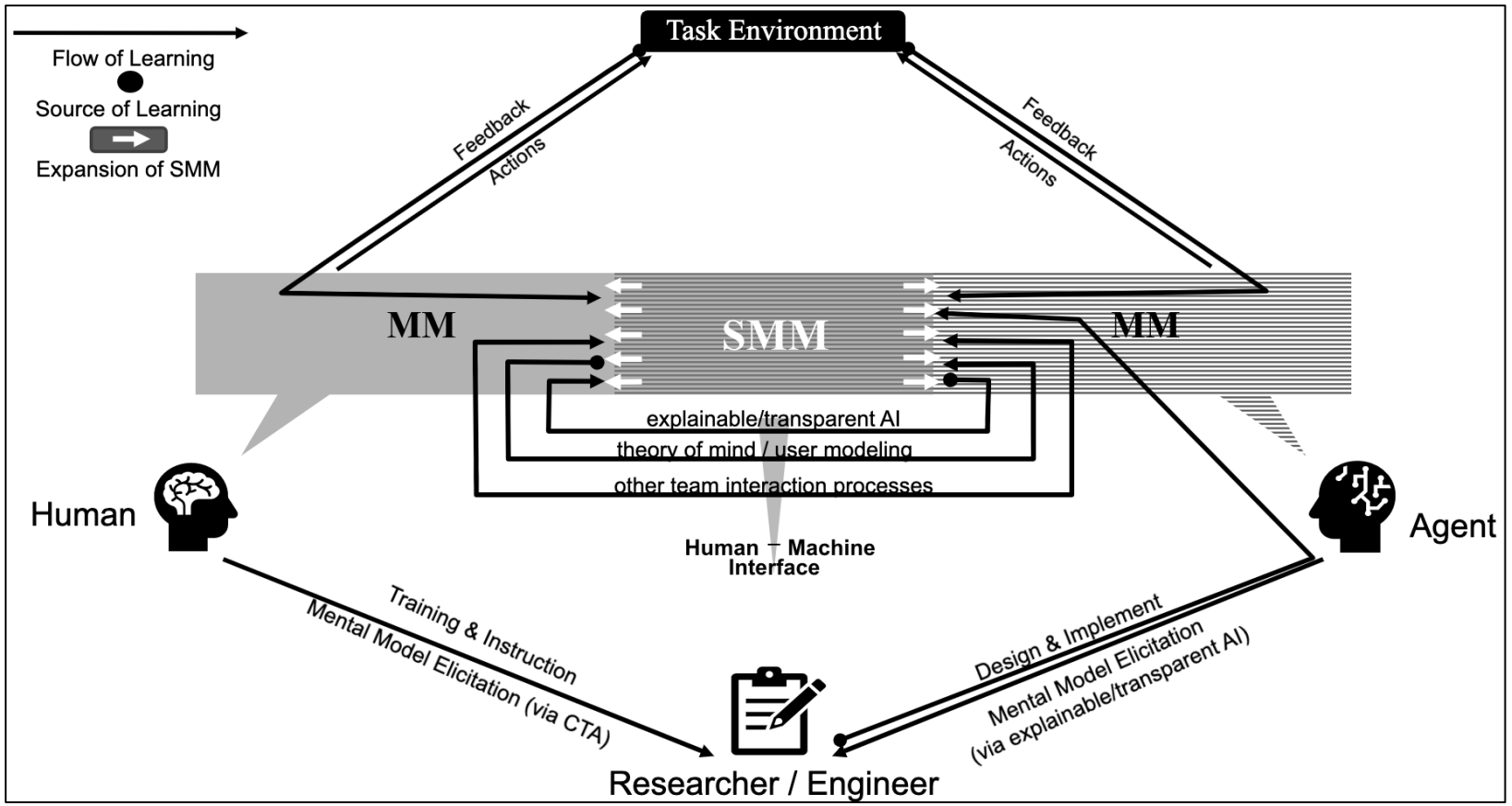
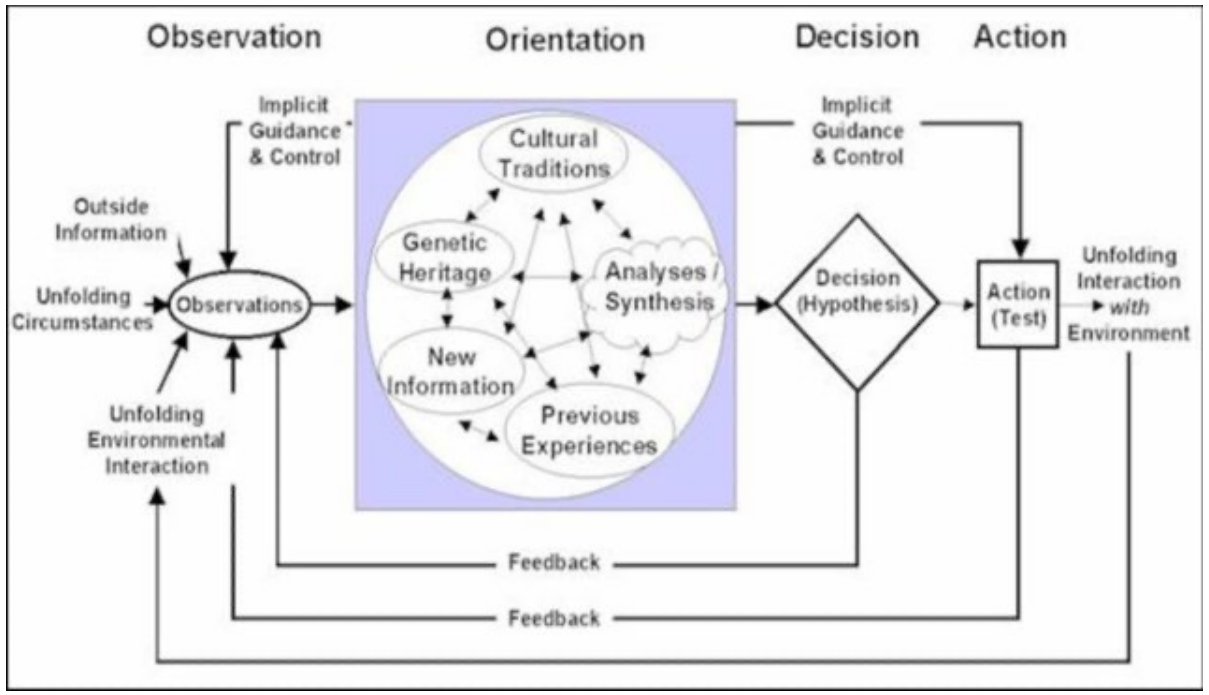


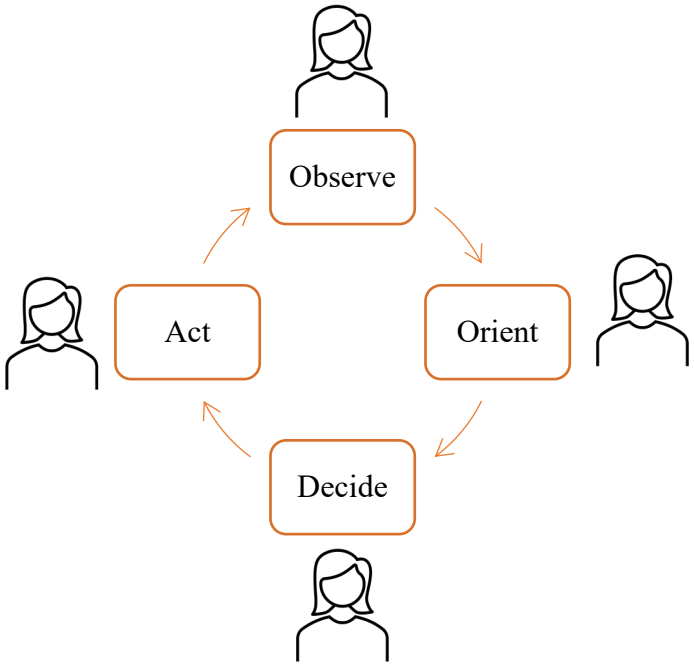
Diagram of an artificial SMM between a human-agent dyad [2]

[1] Orasanu, 1990
[2] Andrews, Lilly, Srivastava, & Feigh, 2022

Ideas and Inspiration #2 – it's not just DM in isolation

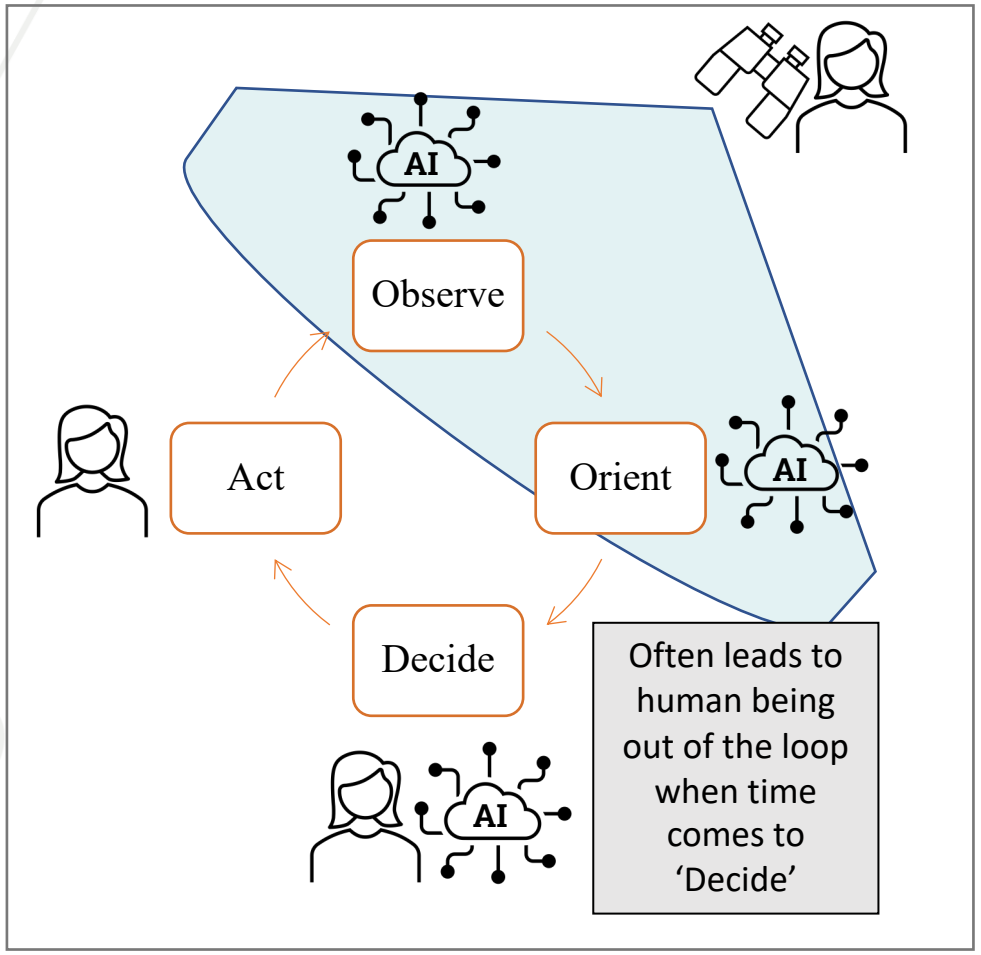


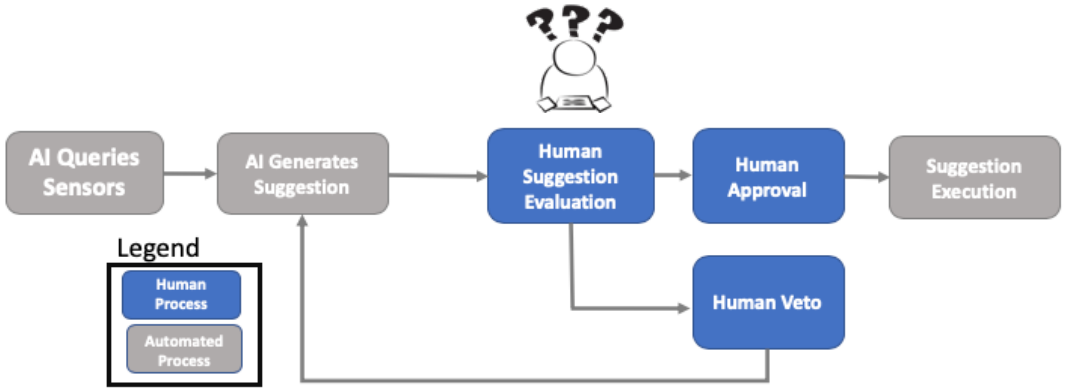
John Boyd's OODA Loop, courtesy of Diane Hendrick



- Decision making does not happen in a vacuum!
- Each cognitive process is dependent on all the others
- Naturalistic decision making (NDM) literature has pointed out that often "decisions" are often trivial when sufficient time and energy has been spent on the Orientation or Judgment activities

Current State of the Art





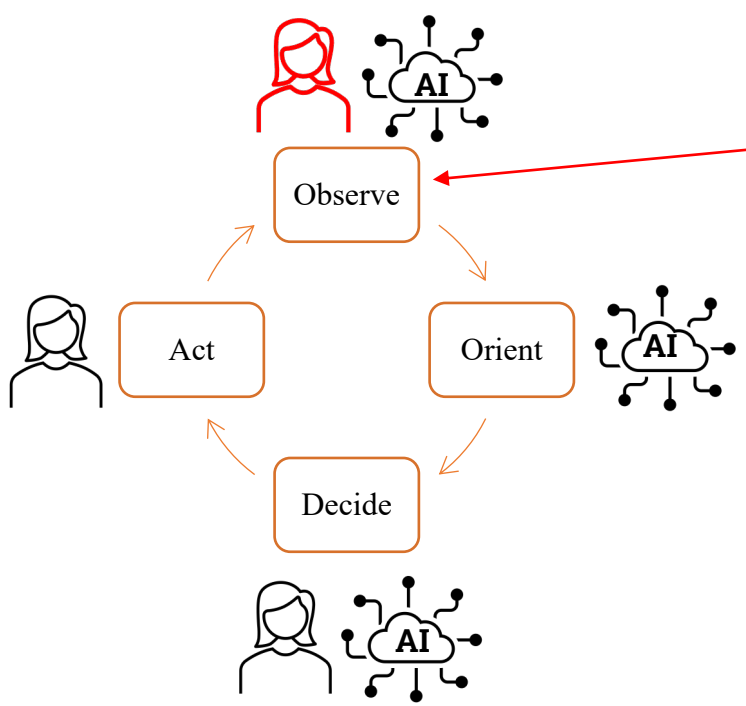
Roles and Responsibilities of the AI:

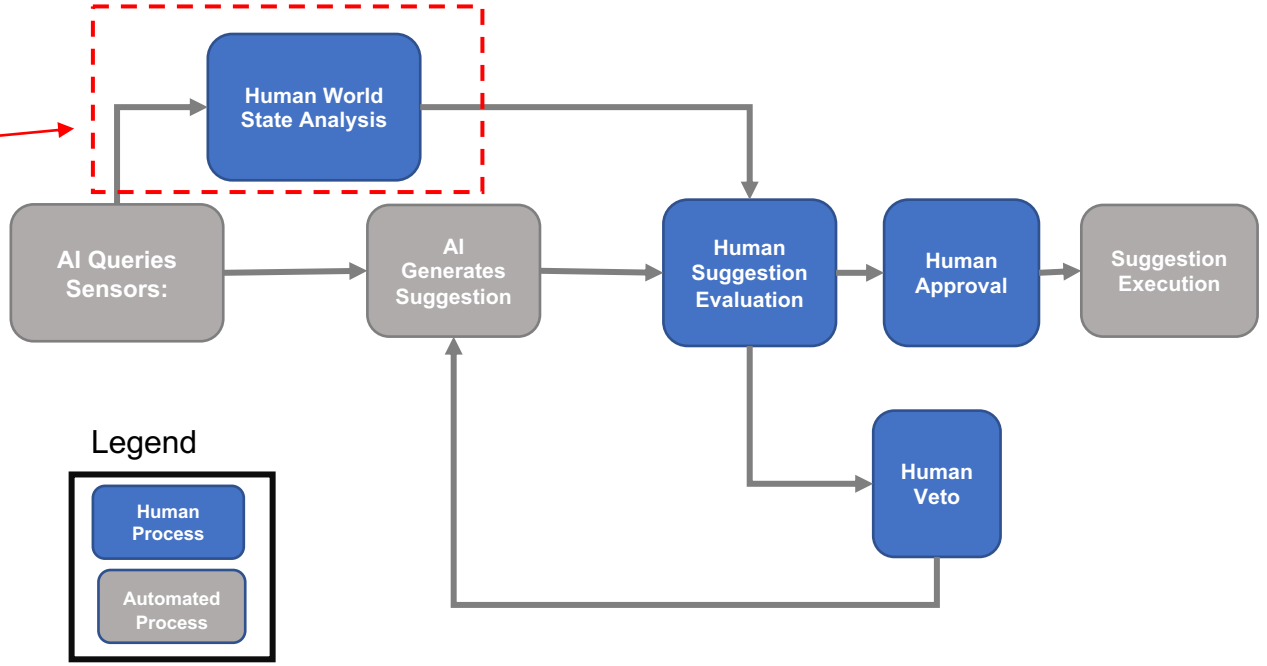
- Observe: Gather relevant information
- Orient: Situate information within mission parameters
- Decision: Suggest a course of action

Roles and Responsibilities of the Human:

- Decision: Commit to a course of action
- Action: Execute the decision
- Supervise/Initiate Observe and Orient

Based on the SMM framework & the emphasis on judgment, **what if we tried to improve the human's understanding of the world state** the automation is basing its suggestion off of?





Roles and Responsibilities of the Human:

- Observe: See the information that the AI is using in its analysis
- Decision: Commit to a course of action
- Action: Execute the decision

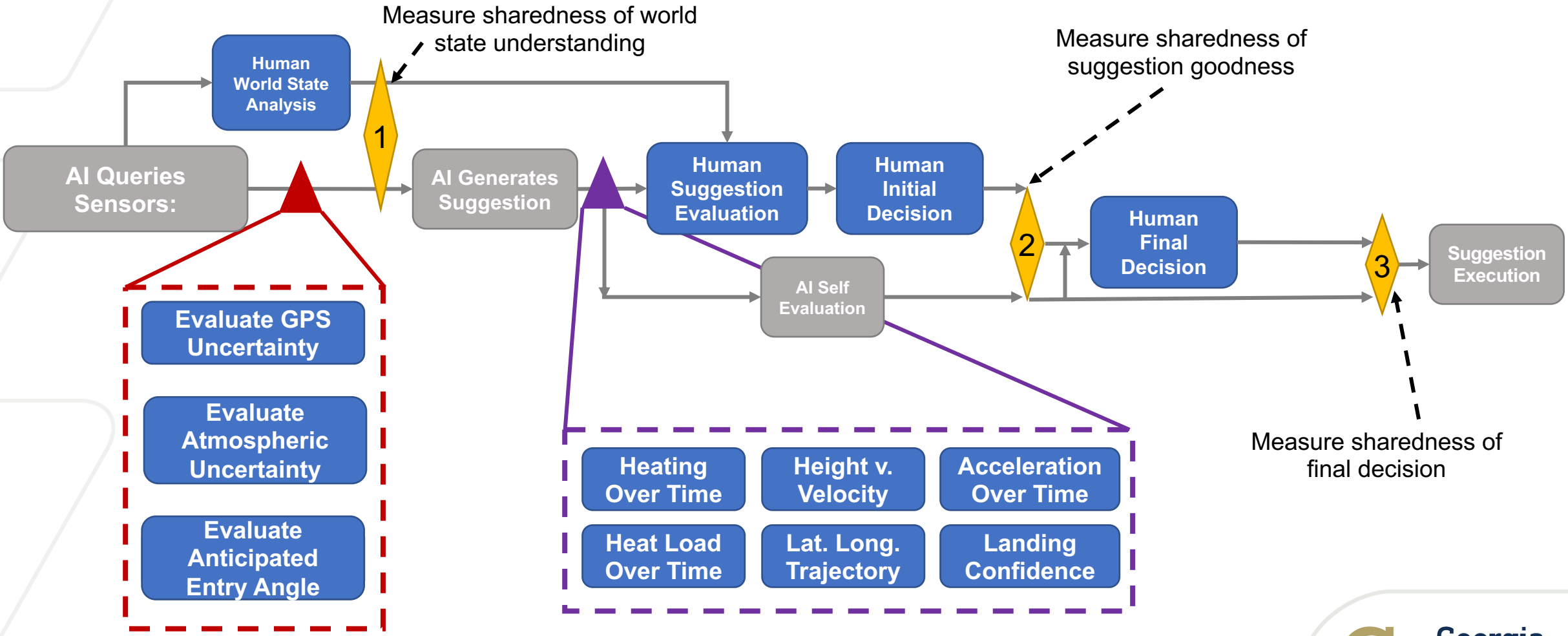
Hypotheses

Emphasizing human **awareness** of the world state (creating shared SA) will

- Improve human robustness to automation over reliance
- Improve team performance
- Improve trust
- Not overly increase workload

Experimental Task Outline/Domain – EDL on Mars

Crew is in orbit around Mars. They need to land a probe, or the mission fails. They have limited information sources. They are working with an automated trajectory planning and analysis system.



Independent Variables of Interest

World State Awareness (3 levels)

- **None** -> task begins after trajectory is presented
- **Observation** → world state is presented, and participants may move on when ready
- **Interaction**→ world state is presented, and participants are asked to answer specific questions.

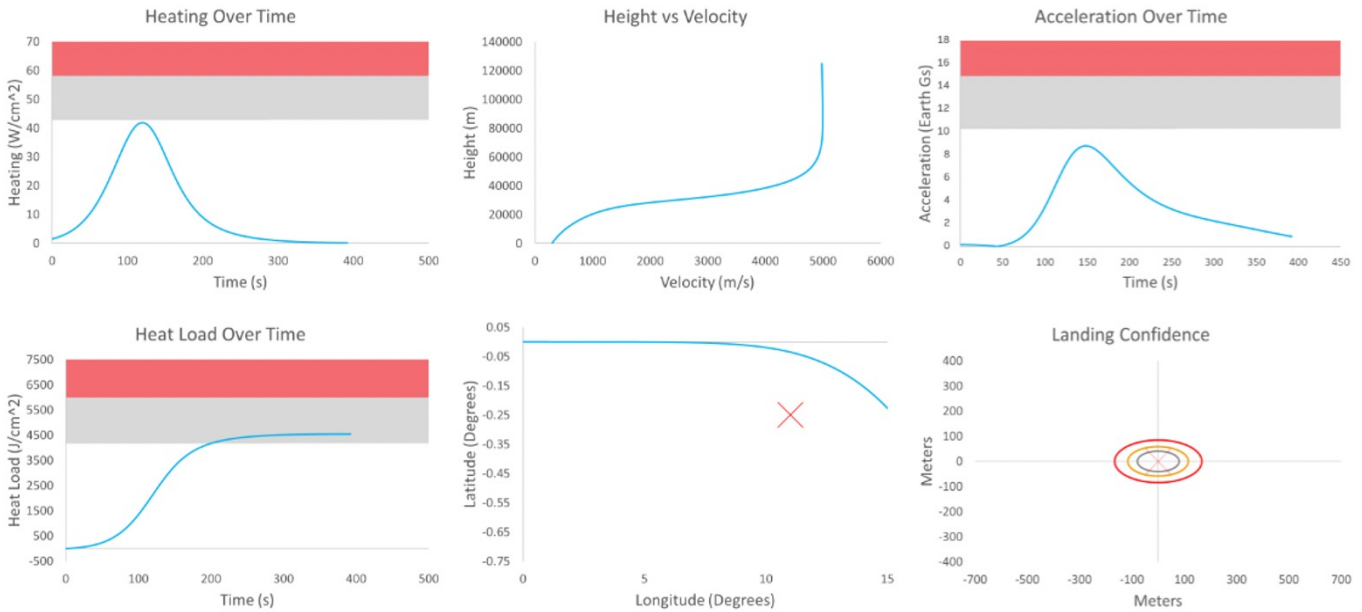
Trajectory Awareness (2 levels)

- **Observation** → trajectory figures of merit are presented, and participants may move on when ready
- **Interaction**→ trajectory figures of merit are presented, and participants are asked to answer specific questions.

Independent Variables of Interest

World State Awareness (3 levels)

- **None** -> task begins after trajectory is presented
- **Observation** → world state is presented, and participants may move on when ready
- **Interaction**→ world state is presented, and participants are asked to answer specific questions.

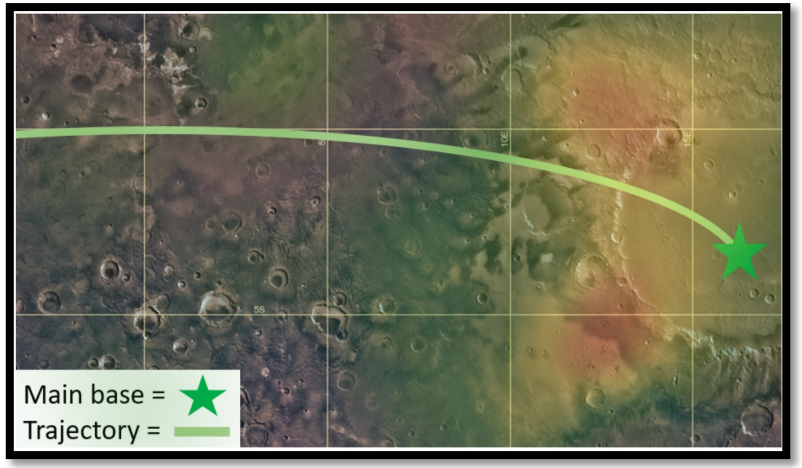


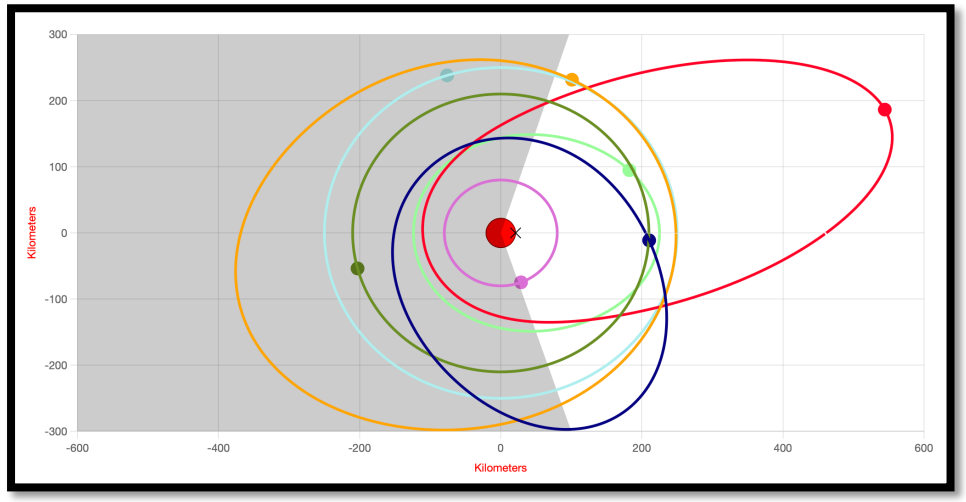
6 Figures of Merit characterizing the trajectory

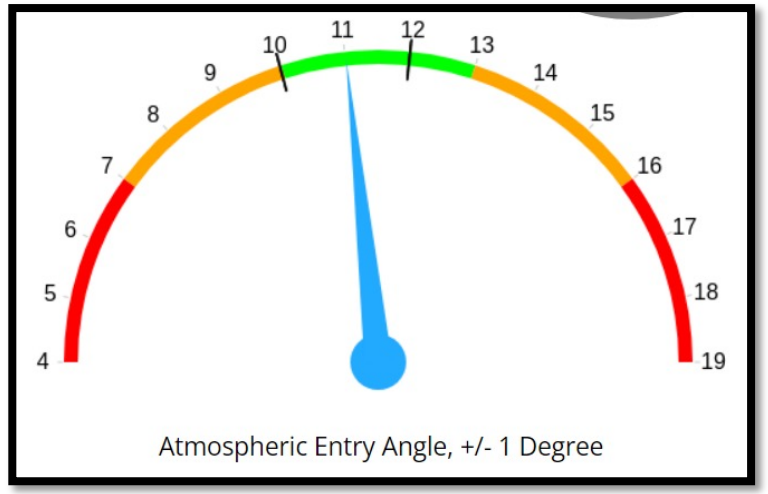
Independent Variables of Interest

World State Awareness (3 levels)

- **None ->** task begins after trajectory is presented
- **Observation →** world state is presented, and participants may move on when ready
- **Interaction→** world state is presented, and participants are asked to answer specific questions.



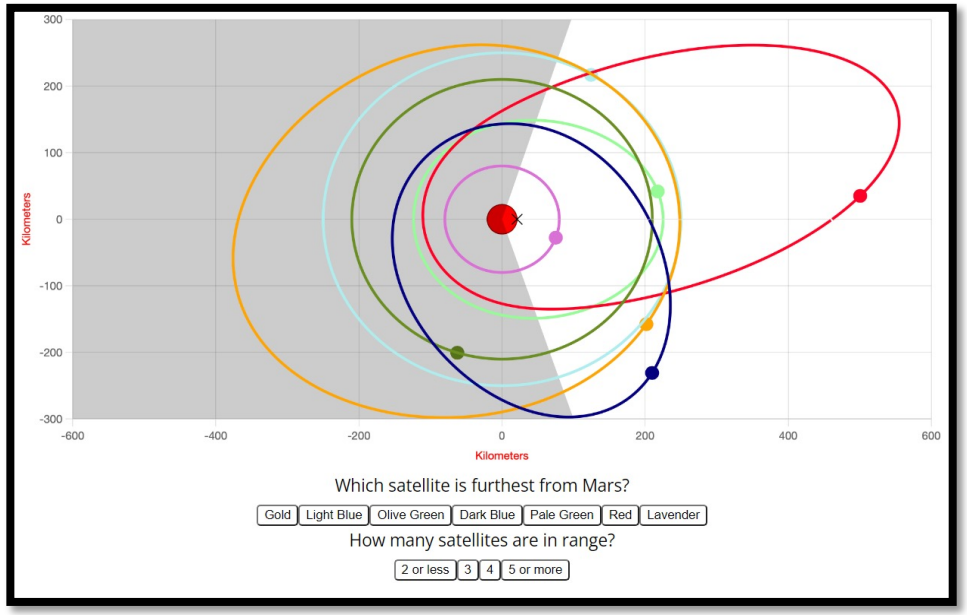


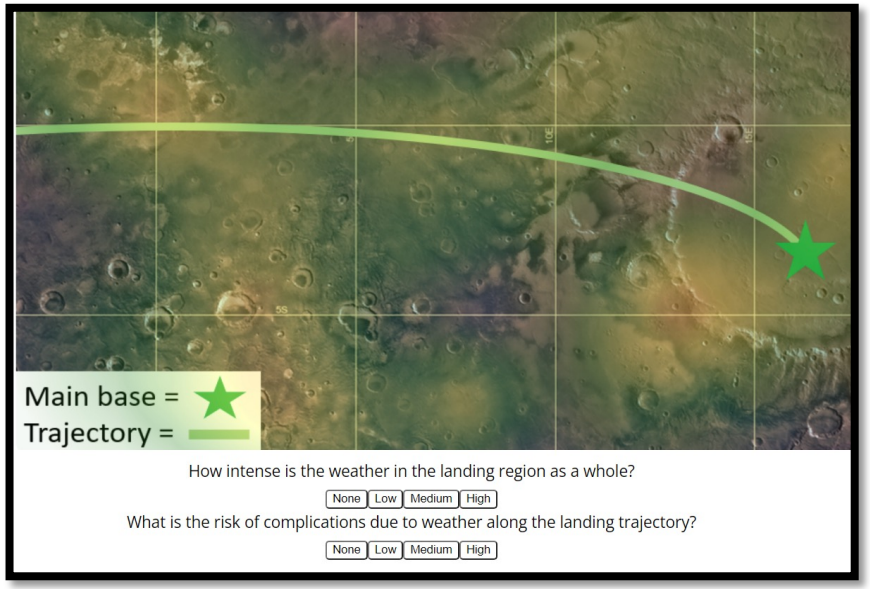


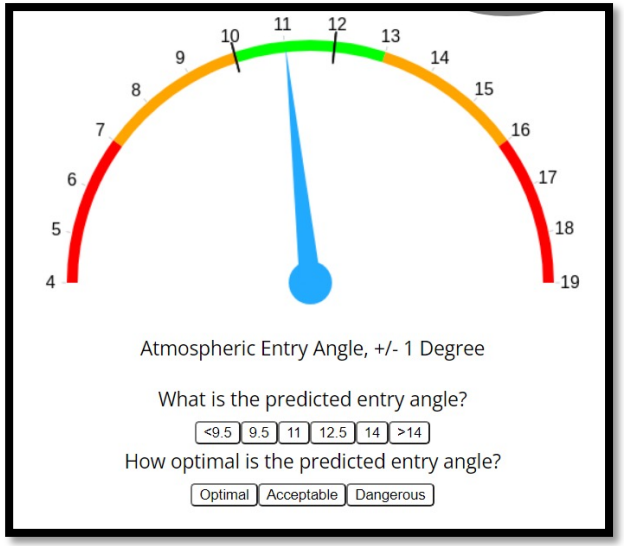
Independent Variables of Interest

World State Awareness (3 levels)

- **None ->** task begins after trajectory is presented
- **Observation →** world state is presented, and participants may move on when ready
- **Interaction→** world state is presented, and participants are asked to answer specific questions.







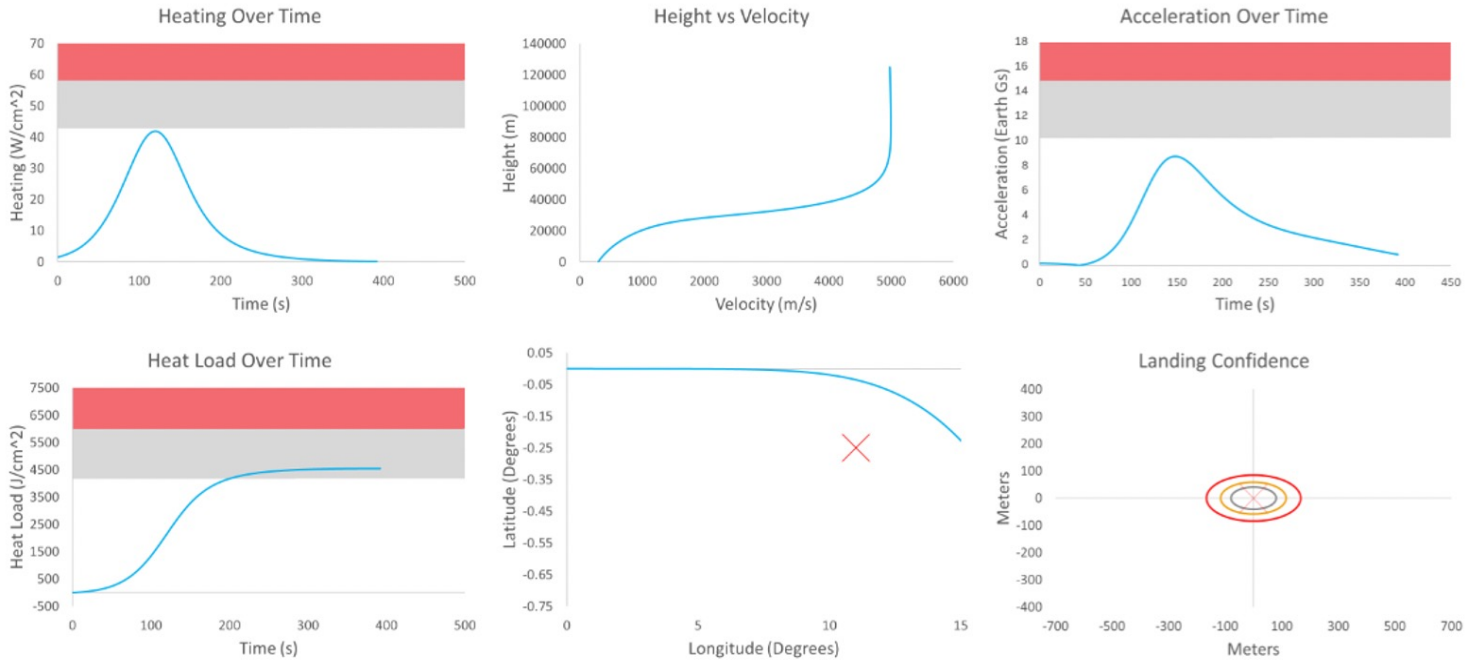
Independent Variables of Interest

World State Awareness (3 levels)

- **None** -> task begins after trajectory is presented
- **Observation** → world state is presented, and participants may move on when ready
- **Interaction**→ world state is presented, and participants are asked to answer specific questions.

Trajectory Awareness (2 levels)

- **Observation** → trajectory figures of merit are presented, and participants may move on when ready
- **Interaction**→ trajectory figures of merit are presented, and participants are asked to answer specific questions.



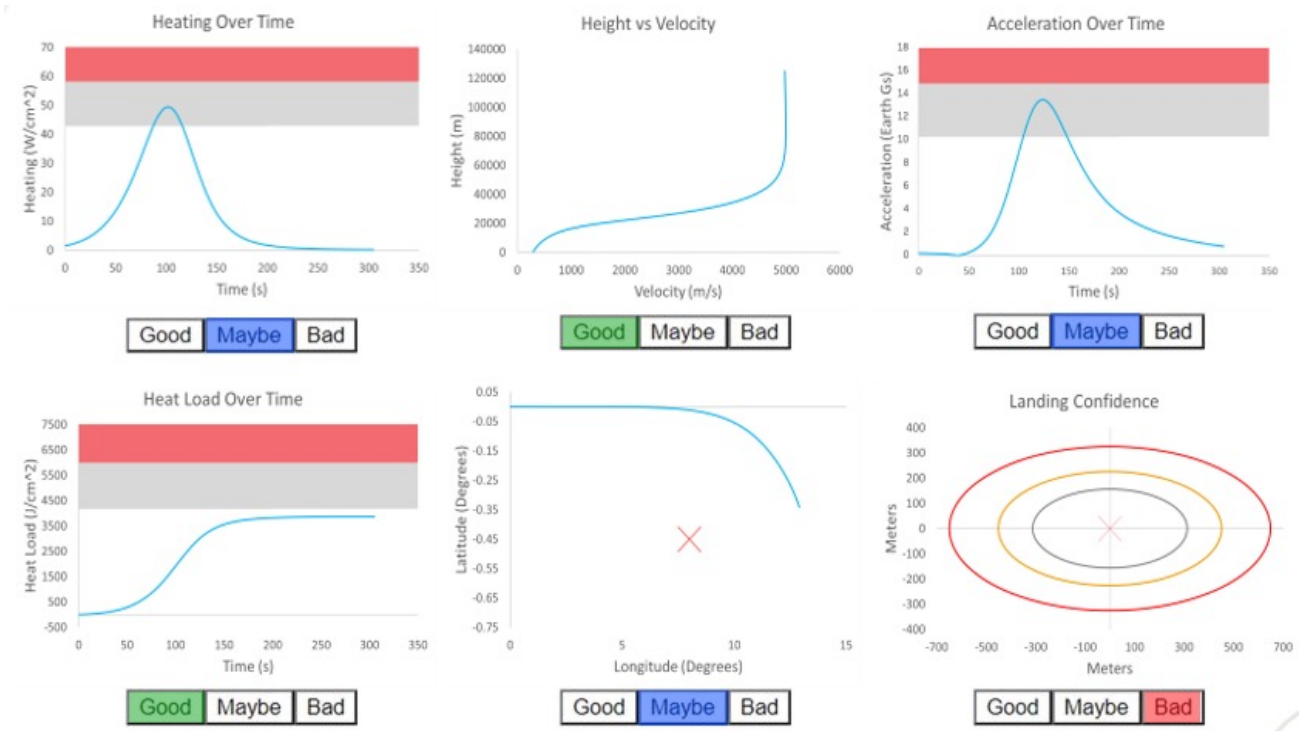
Independent Variables of Interest

World State Awareness (3 levels)

- **None** -> task begins after trajectory is presented
- **Observation** → world state is presented, and participants may move on when ready
- **Interaction**→ world state is presented, and participants are asked to answer specific questions.

Trajectory Awareness (2 levels)

- **Observation** → trajectory figures of merit are presented, and participants may move on when ready
- **Interaction**→ trajectory figures of merit are presented, and participants are asked to answer specific questions.



Independent Variables of Interest

World State Awareness (3 levels)

- **None** -> task begins after trajectory is presented
- **Observation** → world state is presented, and participants may move on when ready
- **Interaction**→ world state is presented, and participants are asked to answer specific questions.

Trajectory Awareness (2 levels)

- **Observation** → trajectory figures of merit are presented, and participants may move on when ready
- **Interaction**→ trajectory figures of merit are presented, and participants are asked to answer specific questions.

Experiment Design: 2 x 3

World State Awareness	Trajectory Awareness	World State Agreement	Trajectory Agreement
None	Obs		X
None	Int		X
Obs	Obs	X	X
Obs	Int	X	X
Int	Obs	X	X
Int	Int	X	X

Procedure

Welcome
- Explain experiment and sign Consent Form

Pre-Experiment Questionnaire
- Participants answer questions that indicate dispositional trust

Training Session
- Participants view training videos and complete 6 training rounds

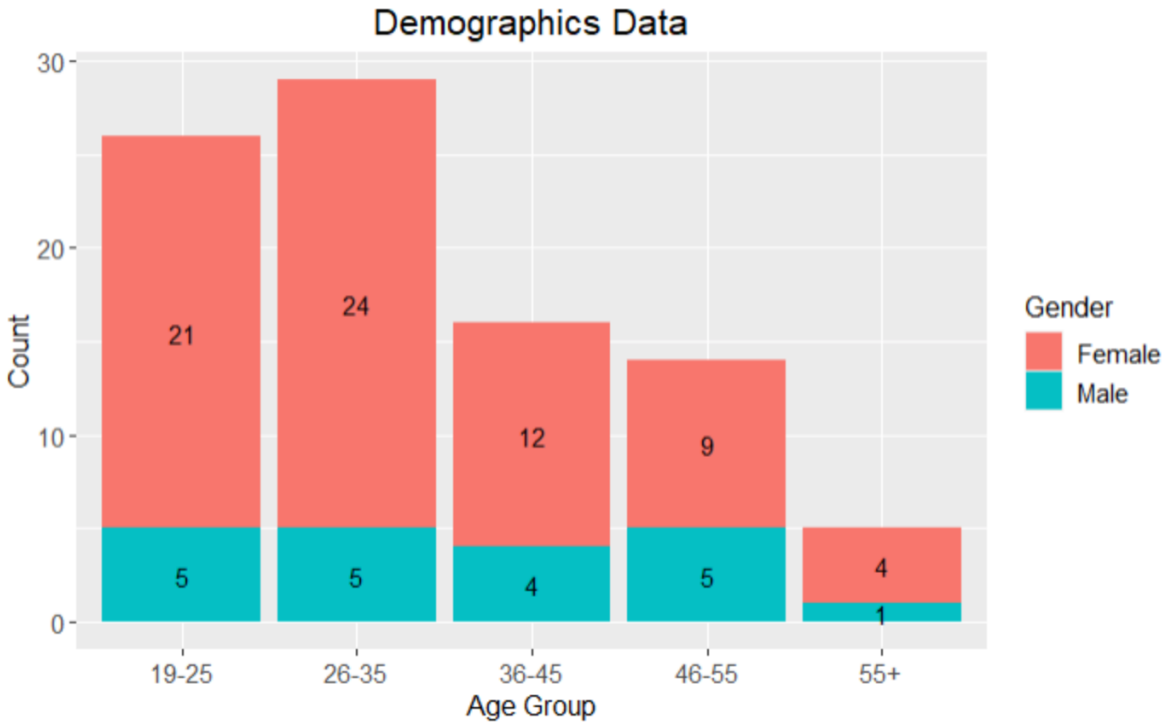
Experiment
- Participants complete 10 rounds

Post-Experiment Questionnaires
- TLX Survey
- Trust survey indicating:

The order in which they saw scenarios was balanced.
Difficulty of scenarios was balanced.

Participants

- Utilized online recruitment platform Prolific
- Experiment 1 (Perfect Automation):
  - Recruited 90 participants (15 per treatment group)
  - Between subject
  - World State Awareness (30 per group)
  - Trajectory Awareness (45 per group)
- Participants were trained to a specific standard
  - Must have been able to correctly answer 5/6 world state awareness questions in the last 3 training scenarios, and get 2/3 scenarios correct.
  - Additionally, we included attention checks and speed warnings



Dependent Variables & Covariates

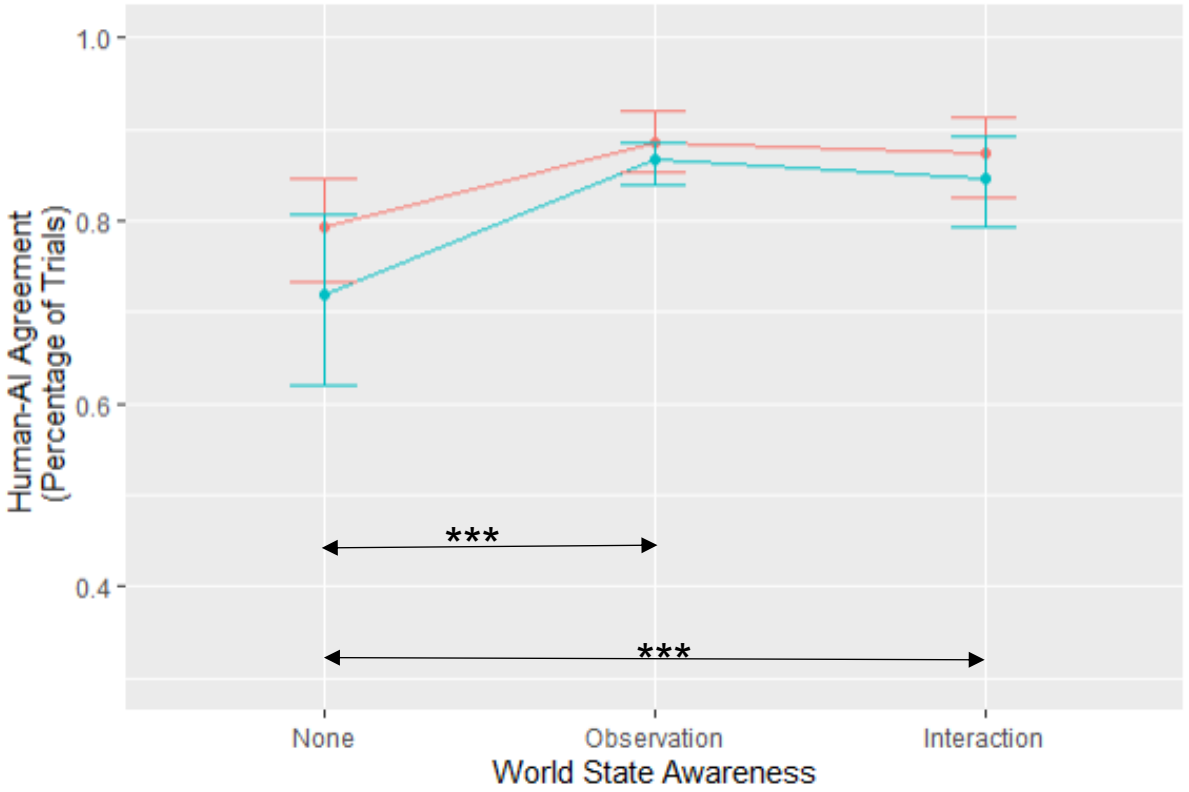
- Agreement % for each stage
  - 0-100% in 10% increments across all 10 data collection trials
- TLX Measures for all data collection trials
- Situational Trust*
  - Measurement of various aspects of trust after all data collection trials
- Dispositional Trust*
  - Measurement of general trust before interaction with experiment

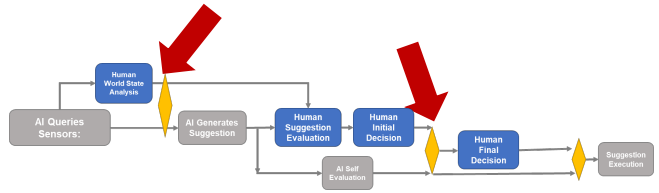
* Measures not presented here due to time constraints

Results

COGNITIVE
ENGINEERING
CENTER

How did world state awareness and trajectory awareness affect shared situation awareness?





Trajectory Awareness

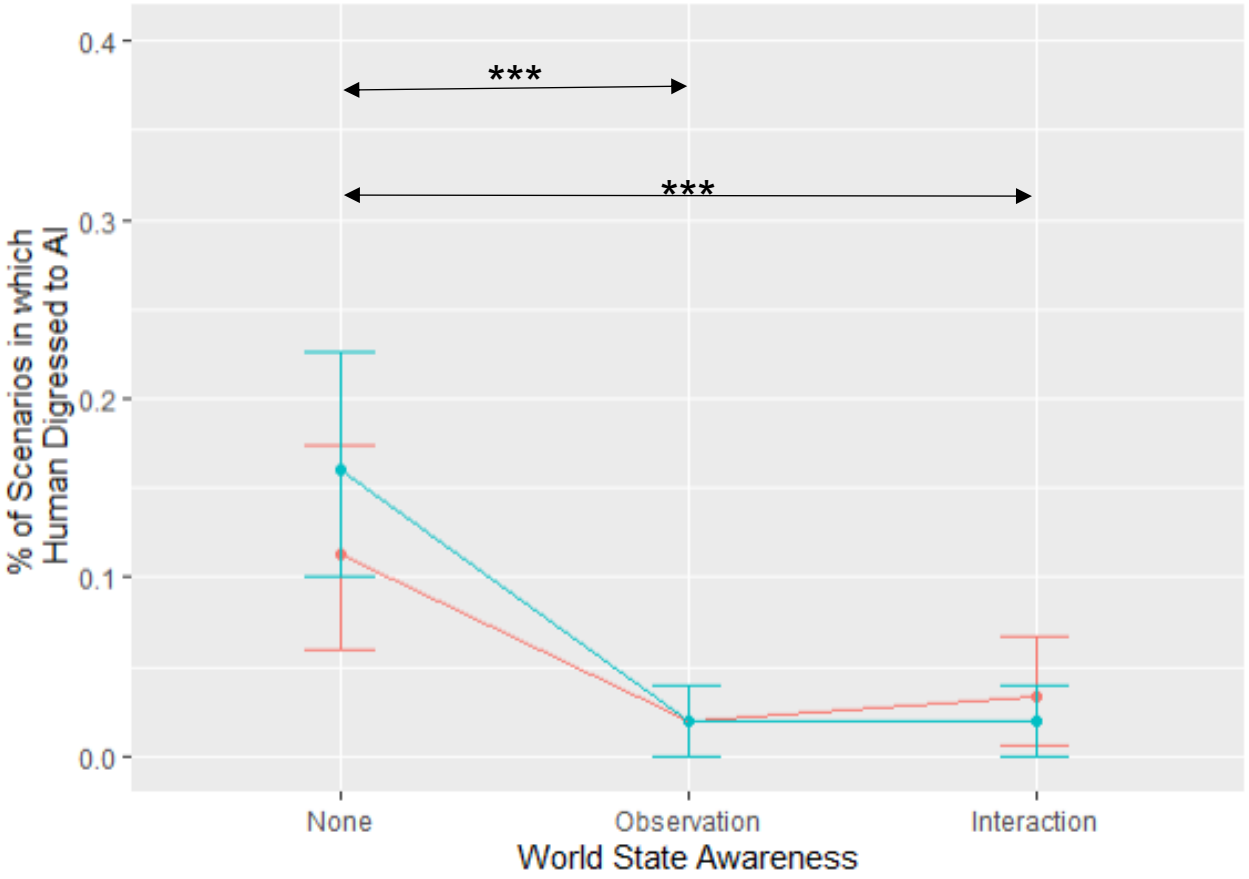
- Observation
- Interaction

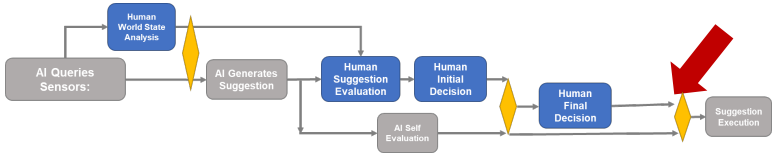
Not statistically significant

World state awareness has significant influence on correct judgement of the black box generated solution

Trajectory awareness (interaction with the black box generated solution) does not have significant influence on correct judgement of the black box generated solution

Who changed their mind after seeing the AI's suggestion?





Trajectory Awareness

- Observation
- Interaction

Not statistically significant

People who did not see world state conditions tended to more readily change their mind to align with the AI (over reliance) the black box generated solution

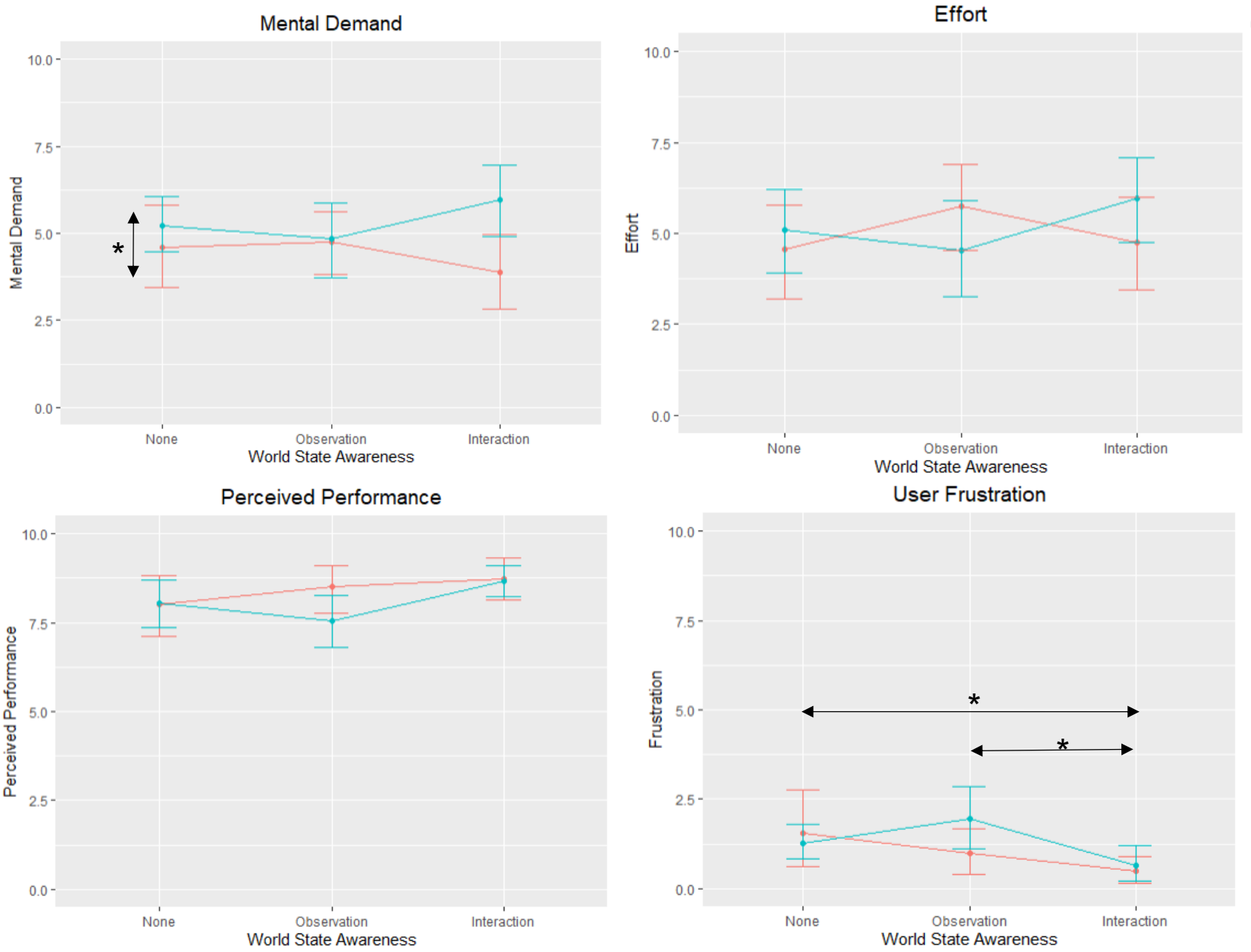
COGNITIVE ENGINEERING CENTER

Georgia Tech

Workload

While interacting with world state conditions required more mental demand and effort on the user's part, it was less frustrating.

More interaction with trajectory charts resulted in worse perceived performance and more mental demand



Key Findings

Putting the human back in the loop by providing world state information improves:

- Human's initial and final judgment of the AI's generated solution
- Human robustness to overreliance

Providing world state information slightly increases mental demand/effort but results in higher confidence and less frustration for the human (aka: ***a better user experience***)

Providing world state information ***accurately calibrates human's trust in AI*** (if the AI is reliable, the human's trust in the AI's capabilities will increase)

Providing interaction as a judgment support technique with world state conditions OR the AI's suggestion improves performance, but ***interacting with both results in weaker performance***

Questions?

Many thanks to Sandia National Labs for funding this work and Paul Schutte for serving as our PM.

This work is our own and does not reflect the official position of Sandia National Labs.

Cognitive Engineering Center

Georgia Tech

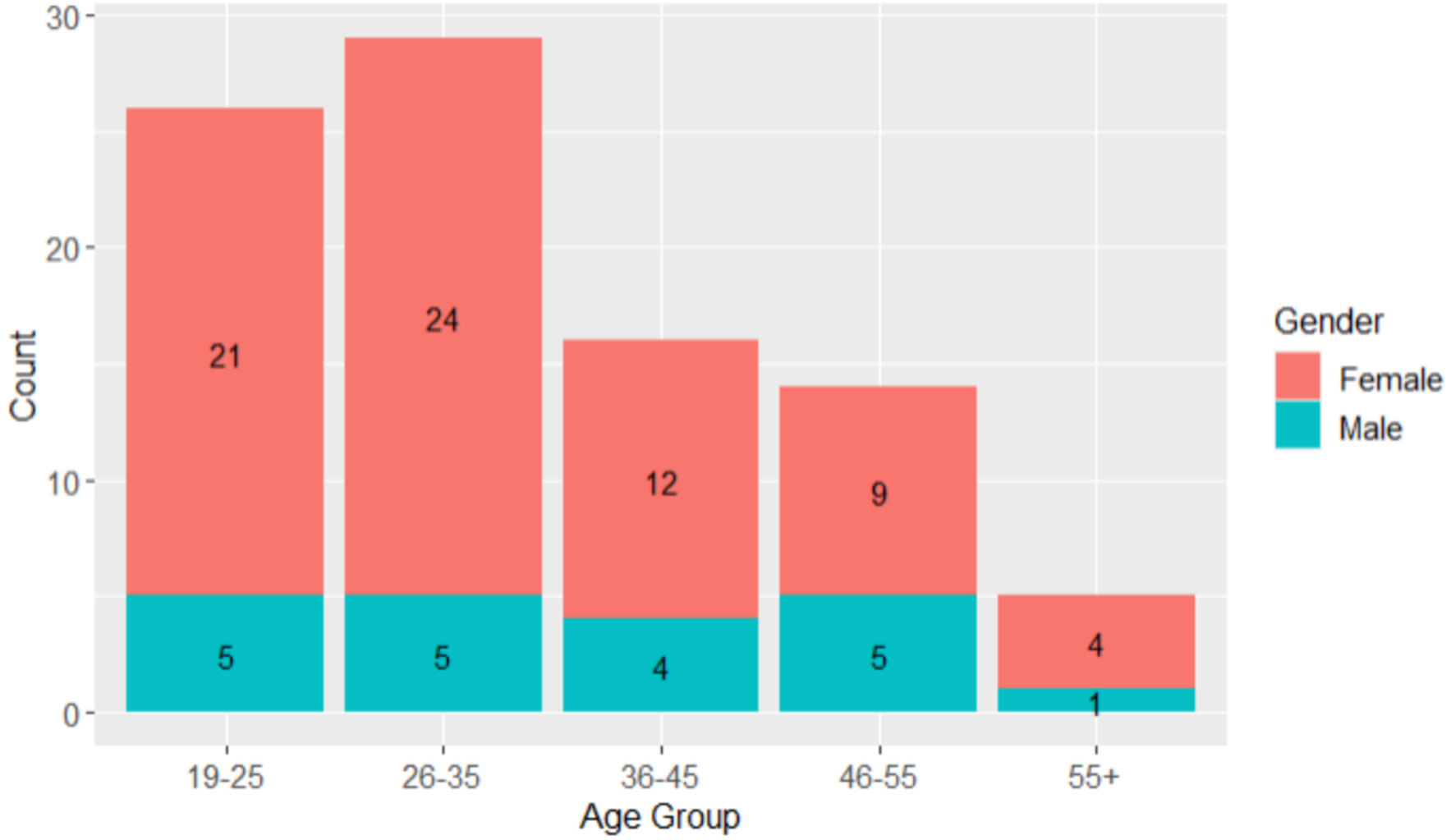
Extra Slides

COGNITIVE
ENGINEERING
CENTER

Data Analysis Methods

- Group means and associated confidence intervals were calculated for each participant across 10 trials
  - World State Sharedness
    - Did the participant correctly assess the world state information? [Risky/Good]
  - Trajectory Sharedness
    - Did the participant make the correct call based on the trajectory information (and world state information if they saw it)? [Execute/Abort]
  - Final Agreement
    - Did the participant make the correct call after seeing the AI's suggestion? [Execute/Abort]
- Two-way Analysis of Variance (ANOVA)
  - Factors: World State Awareness, Trajectory Awareness, Dispositional Trust
  - DV: World State Sharedness, Trajectory Sharedness, Final Agreement
- Multivariate Analysis of Variance (MANOVA)
  - Factors: World State Awareness, Trajectory Awareness
  - DV: TLX Workload Metrics (Mental Demand, Temporal Demand, Physical Demand, User Frustration, Perceived Effort, Perceived Performance)

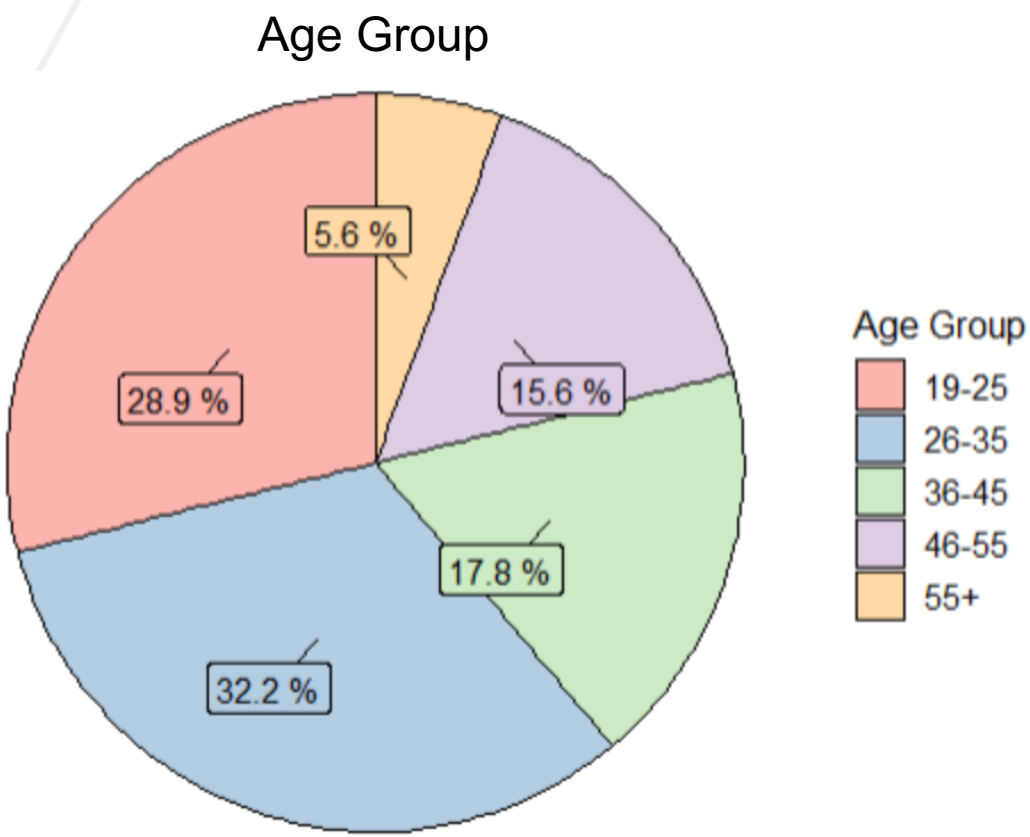
Demographics – 90 Participants

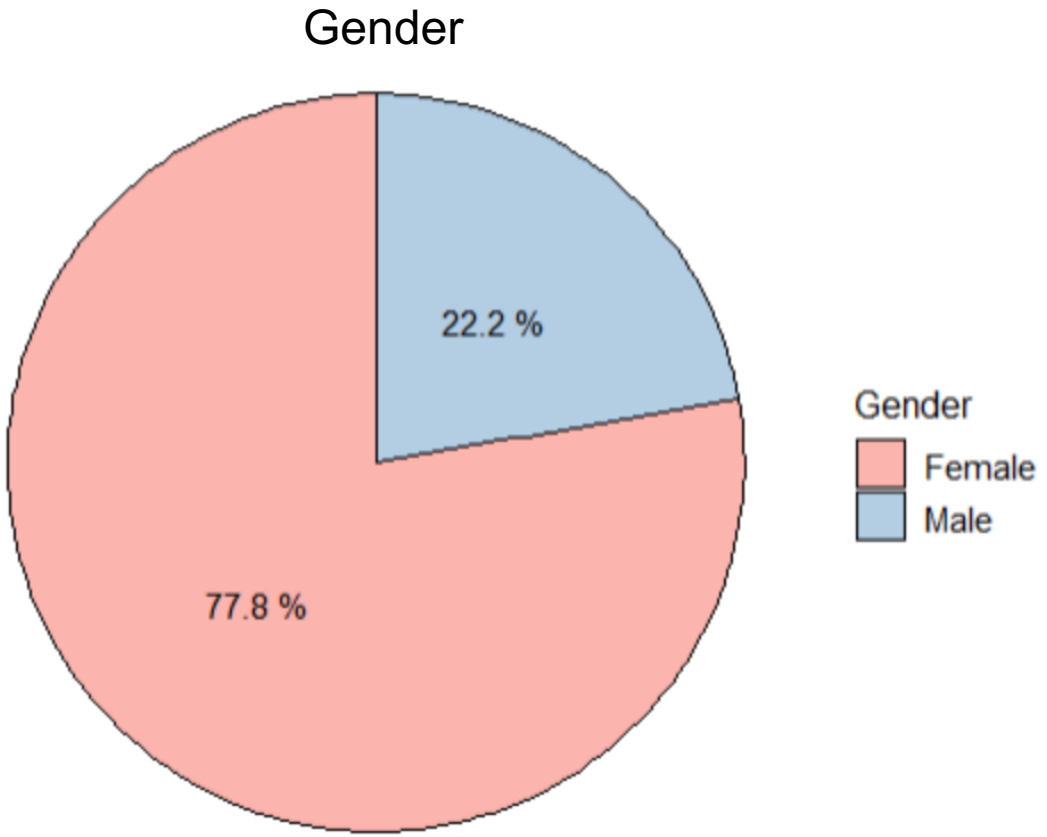


Cognitive Engineering Center

Georgia Tech

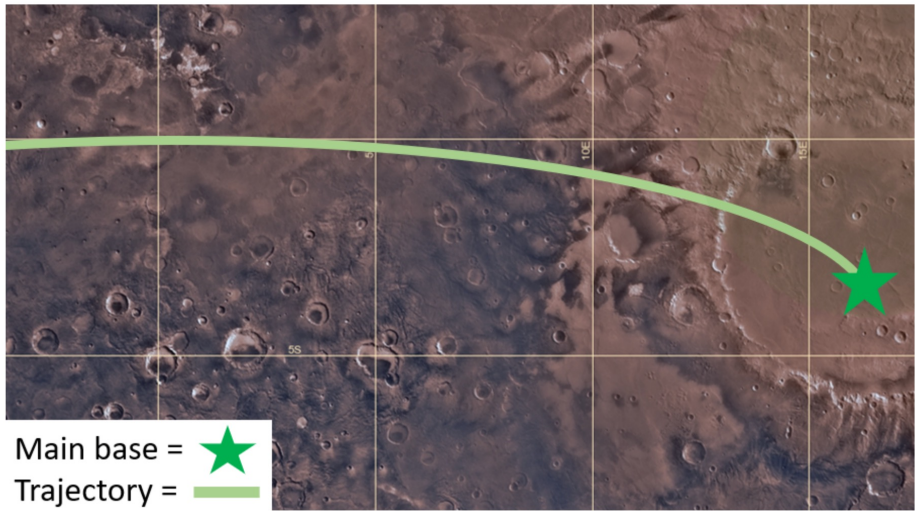
Demographics – 90 Participants



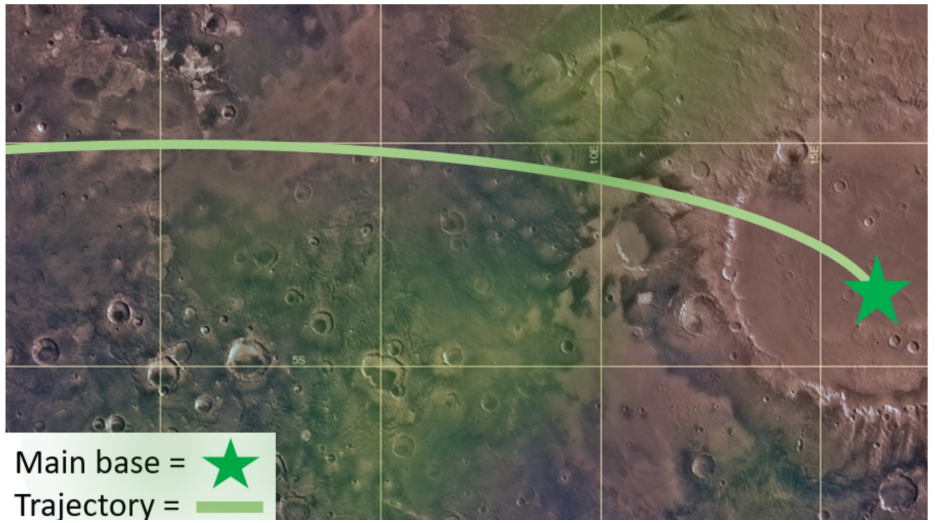


Example Weather Levels

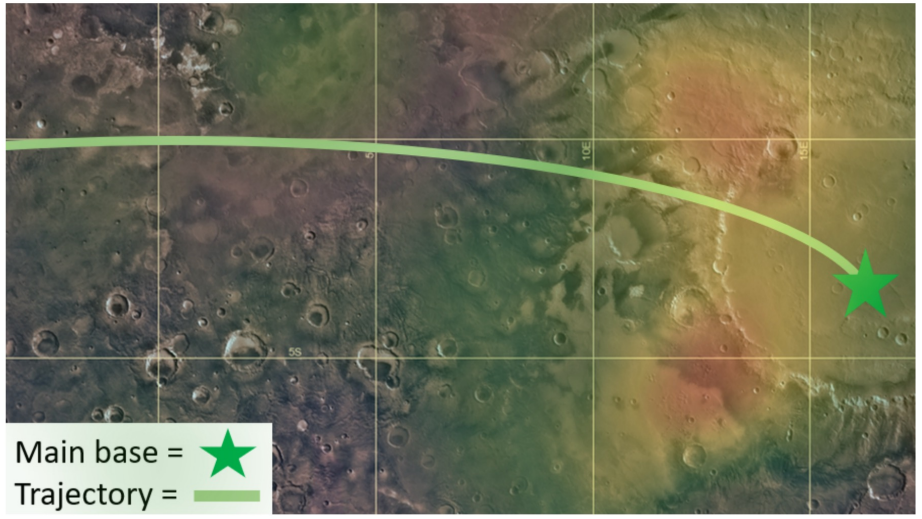
No Risk



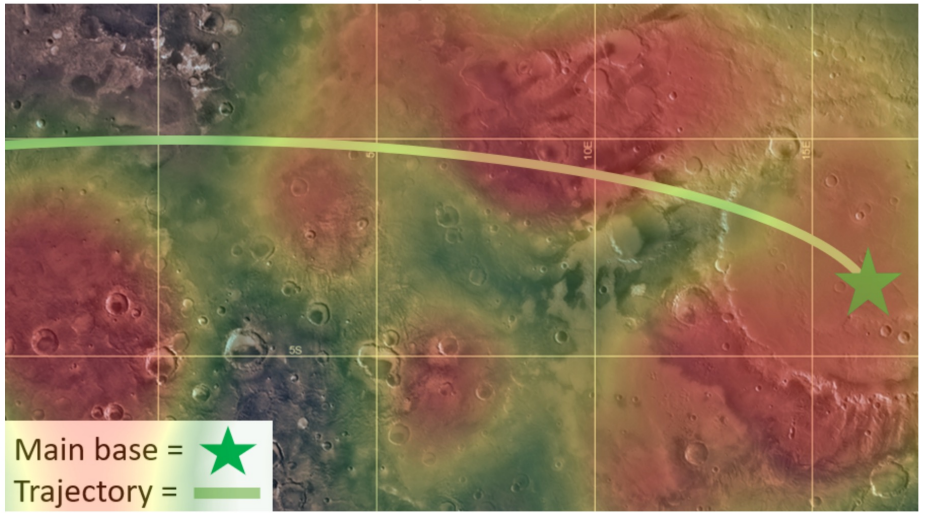
Low Risk



Medium Risk

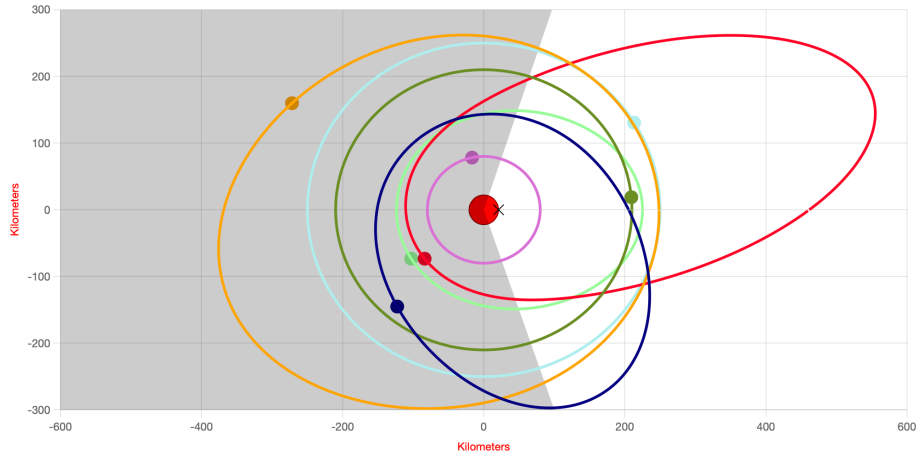


High Risk

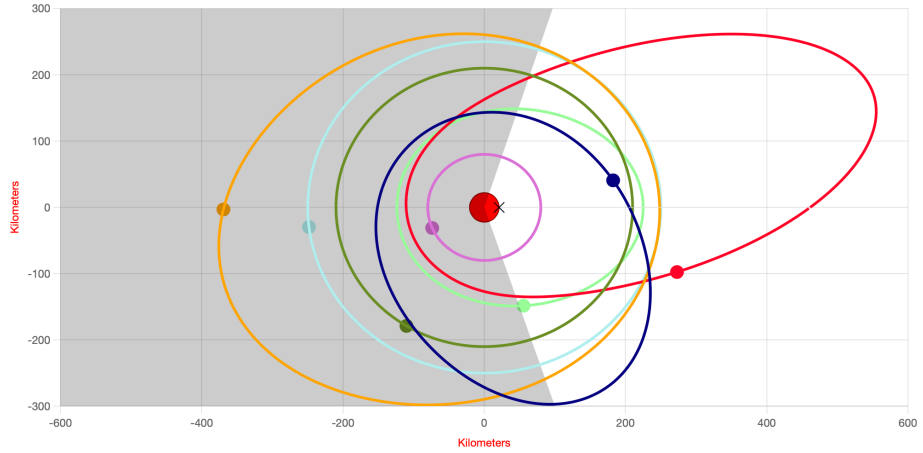


Example GPS Coverage Levels

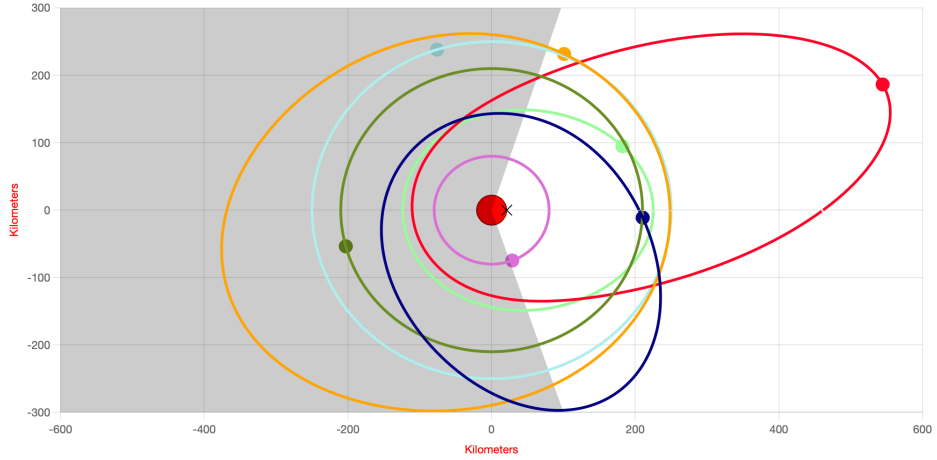
2 Satellites - Bad



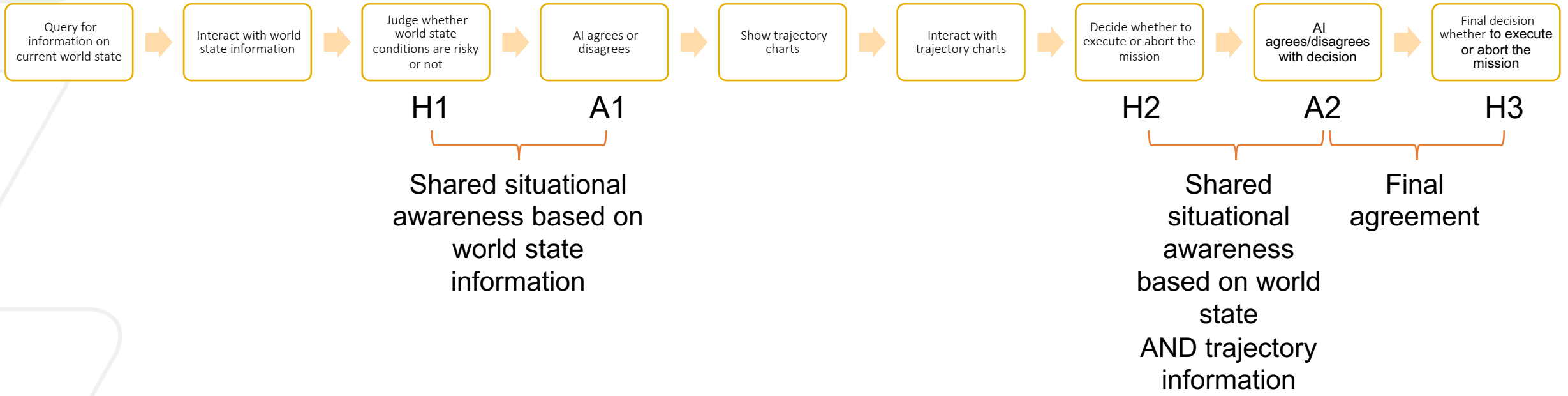
3 Satellites - Okay



5 Satellites – Great!



Most Interactive Treatment



Human AI Teams

- What is the appropriate role for automation/autonomy and conversely humans?
  - Static, dynamic, interchangeable, shifting, etc
- How does having a “learning” system change these challenges?
  - Support individual human needs
    - Novice/expert
    - Preferences
    - Fatigue compensation
  - Adapt to specific vehicles or environments
- How would we allow a system to learn these things?
  - What if they learned wrong?