Impact of Shared Mental Models in Human-AI Decision Making

Ranjani Narayanan, Sarah Walsh

Advised by: Dr. Karen Feigh, Cognitive Engineering Center

Daniel Guggenheim School of Aerospace Engineering

COGNITIVE ENGINEERING CENTER

GT Georgia Tech

Complementarity

Reduced human workload

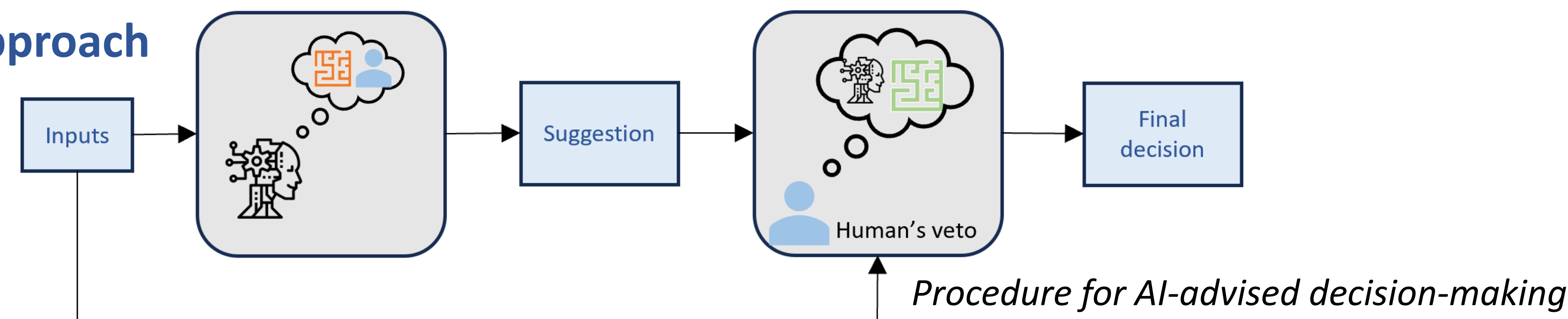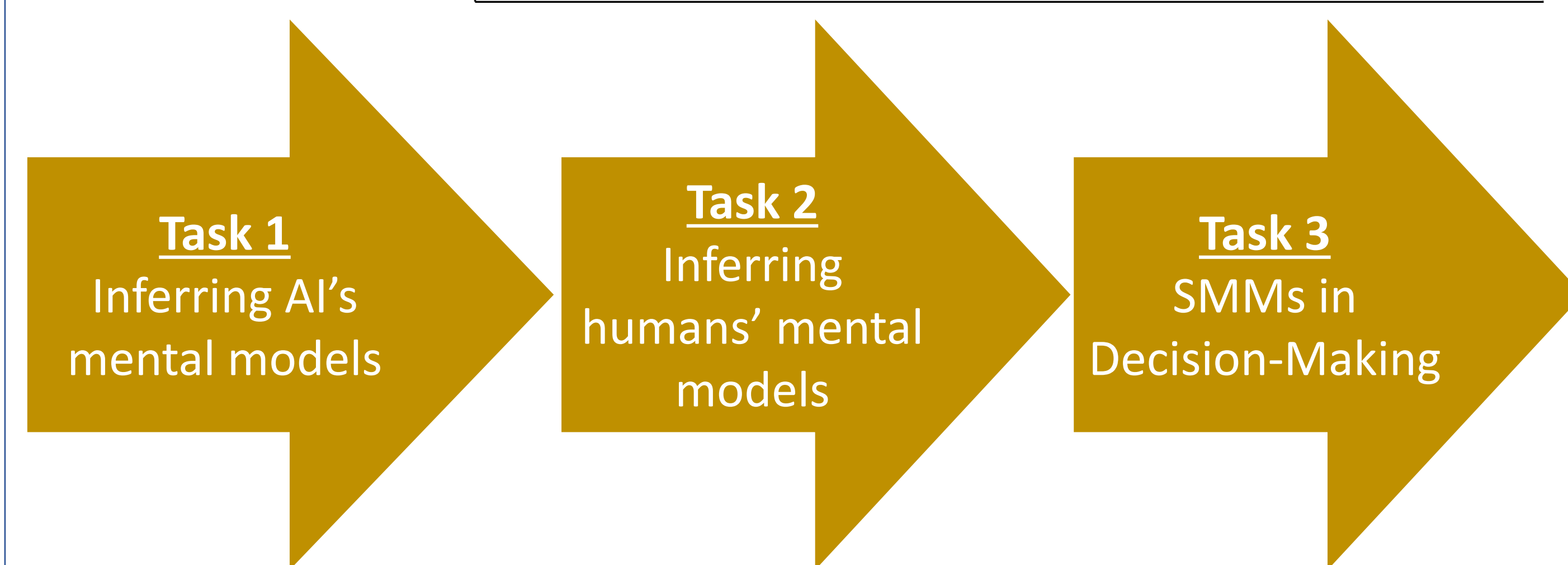Calibrated trust

Fulfill highly complex task objectives

- Humans **understand the world** by constructing working models of it in their minds, called **Mental Models**.
- Human teams are most effective when members of the team utilize a **Shared Mental Model (SMM)** i.e., **shared understanding of goals and actions** through **effective communication and dynamic understanding** of the team and situation.
- Our work explores the importance of **SMMs** in human-AI collaborative teams for decision-making tasks.

Technical Approach

Inputs → Suggestion → Human's veto → Final decision

Procedure for AI-advised decision-making

Task 1 Inferring AI's mental models

Task 2 Inferring humans' mental models

**Task 3** SMMs in Decision-Making

- All experiments are situated around the **AI-advised decision-making** paradigm
- As information is provided, the **AI aid provides a suggestion** to the user
- The **user determines the validity of that suggestion** and makes the final choice
- **Feedback is provided** from the environment for the decision made

Experimental Apparatus

Disaster-relief planning task

CEC CDM Experiment

Data sources: Population, SES, No-go zones, Power Outages, Flooding, Current Storm, Clear

Decision Surface

Tools — Staging site marker: Click on the map to place the marker. Drag the marker to move it. Submit

Toggle panel containing information as heatmaps that can be viewed to determine the best spot to place the resource
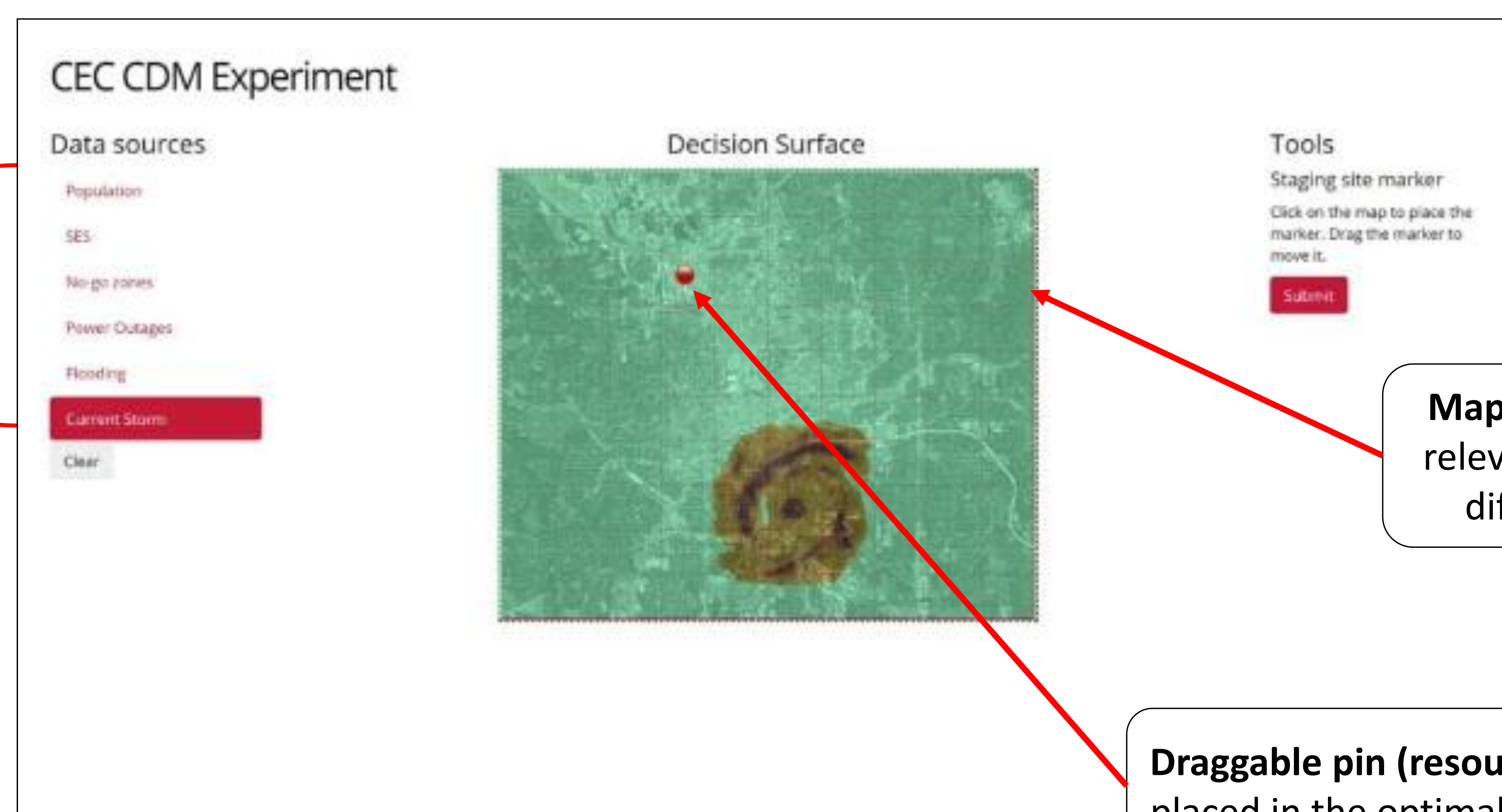
Map overlays showing relevant information of different locations

**Draggable pin (resource)** to be placed in the optimal locations

Participants' goal for this task is to *identify the most optimal spot* to place relevant resources (e.g., food, first aid, etc.) using information from several data sources in anticipation of an oncoming storm.

Inferring AI's 'Mental Model'

The objective is **to identify if humans can discern an AI's knowledge components** by identifying its error boundaries of varying complexity and severity. We also seek **to identify if humans can overcome the AI's lack in performance** and thus, **calibrate their reliance** on the AI decision aid.

Findings: 1. AI error boundaries are identifiable when they are simplistic.

2. Users who improve team performance are accurate in identifying errors as opposed to those who degrade team performance.

3. User agreement with the AI and their perception of its capabilities are reduced as the AI errs, regardless of the error complexity and severity.
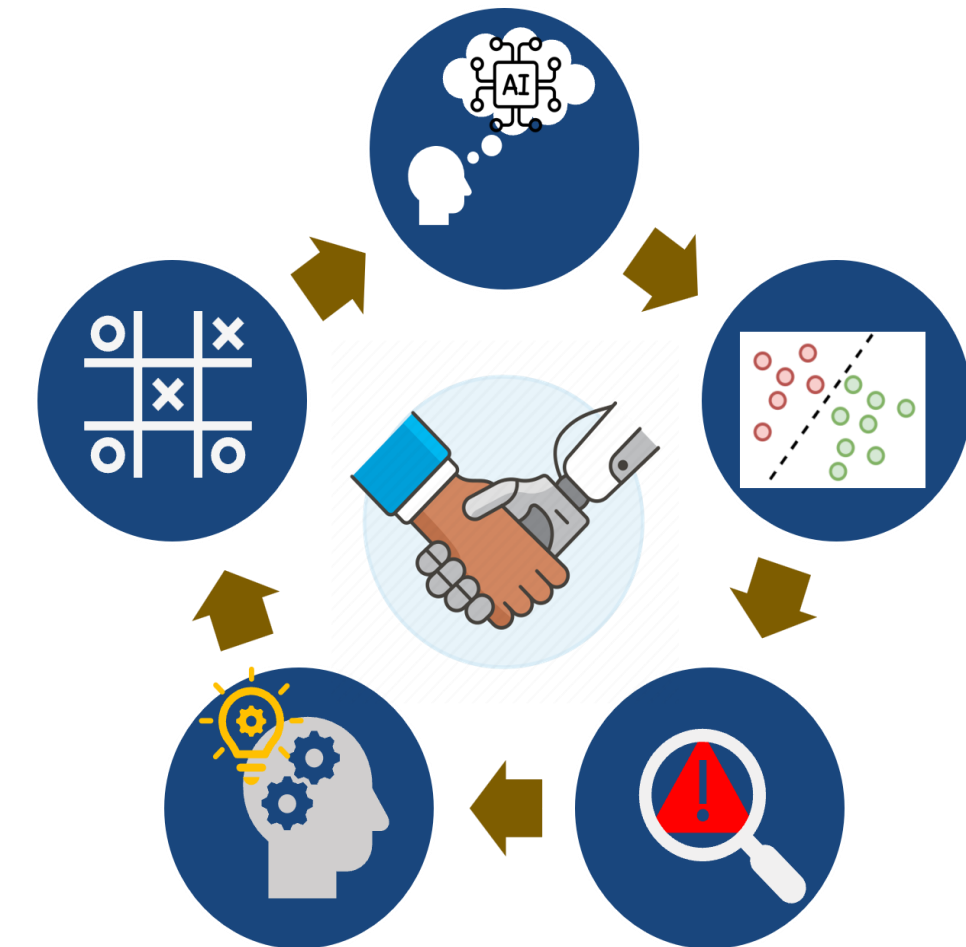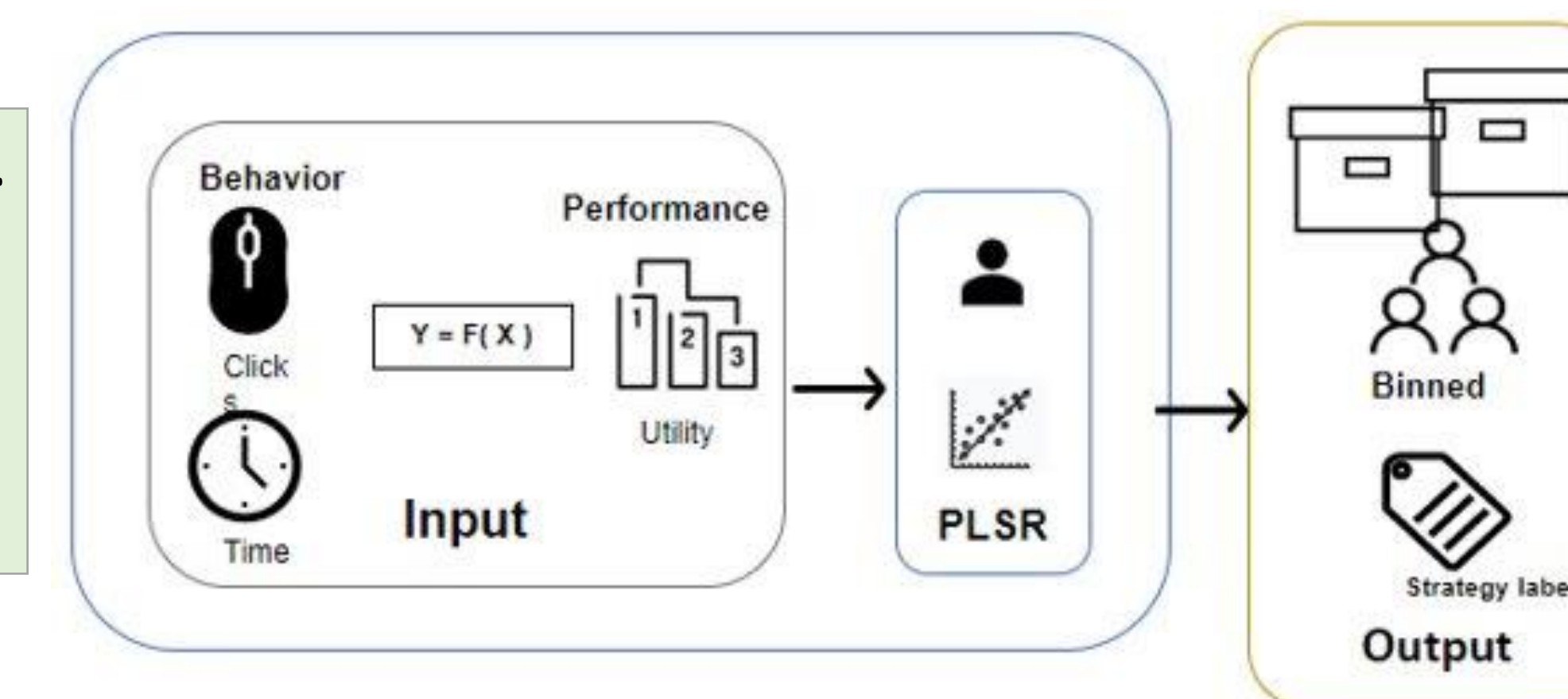
Inferring Human's Mental Model

The objective is **to identify if algorithms may identify humans' mental models of the task** in the form of decision strategies by observing their interaction and behavior without explicitly acquiring any other information from the decision-maker.

Findings: 1. User strategies can be inferred from their behavior.
2. Most users are imperfect decision-makers.
3. The best performers adapt to a suitable strategy, while poorer performers are less adaptive and risk averse.
4. User strategies become predictable with task familiarity.

Behavior — Click, Time — Performance — $Y = F(X)$, Utility — Input — PLSR — Binned — Strategy labels — Output

Shared Mental Models in Decision Making

The objective is **to identify the impact of humans and AI possessing mental models of each other** and if having a **shared understanding of the task** impacts the **performance and workload experienced** by decision-makers.

Findings: 1. Users with a mental model of the AI compensate for performance reduction due to errors.
2. Teams with AI having a model of the user use lesser time and effort to maintain performance levels.
3. Teams with agents having a model of each other perform better than teams with no models and one-way models (only one agent has a model of the other).
4. Lastly, teams with complete SMMs outperform all other teams.

Two way model — Complete SMM — One way model
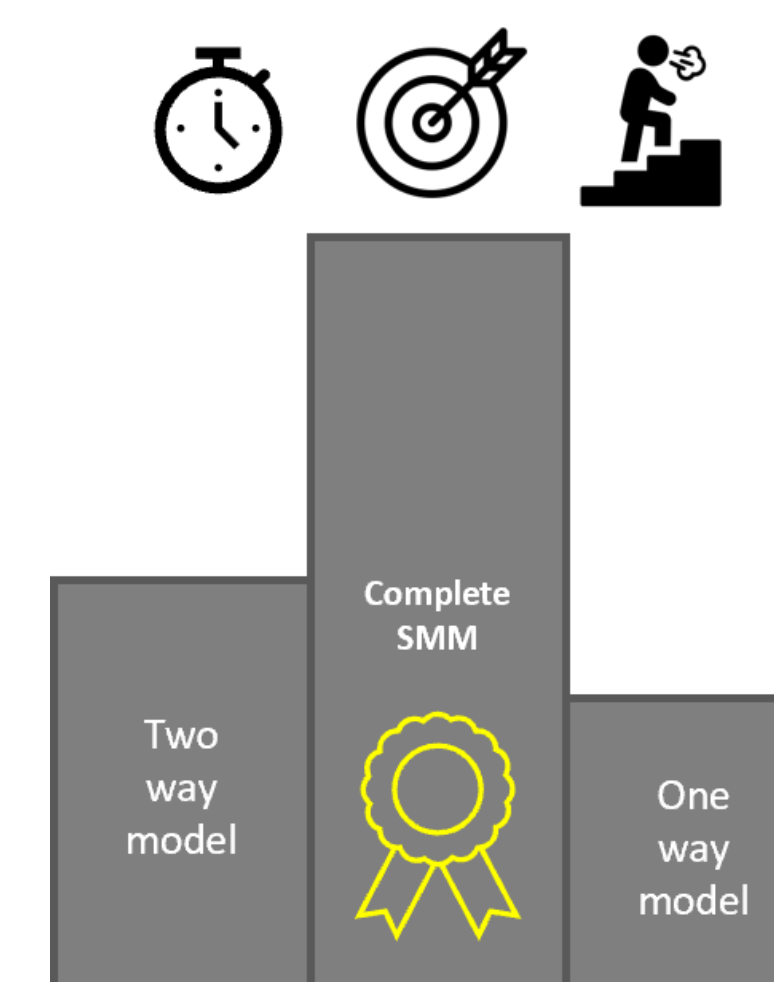
Discussion

Establishing Shared Mental Models in Human-AI teams has improved the quality of decisions made.

- As AI is being increasingly used in decision-making in a variety of it becomes **increasingly important for humans to understand the AI's strengths, weaknesses and tendencies, to be able to use them effectively**.
- **Decision-support tools must support the human by understanding their cognitive demands**, so that they can improve decisions and **achieve greater performance than any human or AI alone**.
- Research efforts should focus on identifying the humans' cognitive states while interacting with AI agents to enable formation of better SMMs for effective collaboration.

Publications: [1] Narayanan, R., Walsh, S. E., & Feigh, K. M. (2023). Development of Mental Models in Decision-Making Tasks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 67(1), 767-773. https://doi.org/10.1177/21695067231192195
[2] Narayanan, R., & Feigh, K. M. (in preparation) Mental Model Inference through Error Identification and Implications in Hybrid Teaming for Decision Making. *ACM Transactions on Compute Human Interaction*
[3] Walsh, S. E., & Feigh, K. M. (2022). Understanding Human Decision Processes: Inferring Decision Strategies From Behavioral Data. Journal of Cognitive Engineering and Decision Making, 16(4), 301-325. https://doi.org/10.1177/15553434221122899
[4] Walsh, S. E., Narayanan, R., & Feigh, K. M. (in preparation) The role of Shared Mental Models in AI-advised Decision Support. Journal of Cognitive Engineering and Decision Making.