

# Predicting and Understanding Human Action Decisions: Insights from Large Language Models and Cognitive Instance-Based Learning

Thuy Ngoc Nguyen<sup>1</sup>, Kasturi Jamale<sup>1</sup>, Cleotilde Gonzalez<sup>2</sup>

<sup>1</sup> Department of Computer Science, University of Dayton, OH, 45469, USA

<sup>2</sup> Department of Social and Decision Sciences, Carnegie Mellon University, PA, 15213, USA

ngoc.nguyen@udayton.edu, jamalek1@udayton.edu, coty@cmu.edu

## Abstract

Large Language Models (LLMs) excel in tasks from translation to complex reasoning. For AI systems to help effectively, understanding and predicting human behavior and biases is essential. However, it remains an open question whether LLMs can achieve this goal. This paper addresses this gap by leveraging the reasoning and generative capabilities of LLMs to predict human behavior in two sequential decision-making tasks. These tasks involve balancing between exploratory and exploitative actions and handling delayed feedback, which is essential for simulating real-life decision processes. We compare the performance of LLMs with a cognitive instance-based learning (IBL) model, which imitates human experiential decision-making. Our findings indicate that LLMs excel at rapidly incorporating feedback to enhance prediction accuracy. In contrast, the IBL model better accounts for human exploratory behaviors and effectively captures loss aversion bias — the tendency to choose a sub-optimal goal with fewer step-cost penalties rather than exploring to find the optimal choice, even with limited experience. The results highlight the benefits of integrating LLMs with cognitive architectures, suggesting that this synergy could enhance the modeling and understanding of complex human decision-making patterns.

## Introduction

Understanding and predicting human behavior in decision-making settings is crucial for developing AI systems that can effectively collaborate with people and help them make informed decisions and avoid cognitive biases and limitations (Hoffman, Bhattacharjee, and Nikolaidis 2023; Bansal et al. 2019; Zhang et al. 2021; Rastogi et al. 2023). One common approach to predicting individual behavior is using machine learning techniques to model their decision-making processes based on past behaviors. These techniques include imitation learning (e.g., behavior cloning (Torabi, Warnell, and Stone 2018)) and machine theory of mind (ToM) (i.e., ability to reason about the mental states of others, including) (Abbeel and Ng 2004), Bayesian ToM (Baker et al. 2017), or neural networks (Rabinowitz et al. 2018)). However, these methods often require extensive training datasets and struggle to model human decision-making accurately with limited samples.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

The recent rise of large language models (LLMs) such as ChatGPT (Ouyang et al. 2022), PaLM (Chowdhery et al. 2023), and LLaMA (Touvron et al. 2023) has demonstrated their remarkable capabilities in semantic understanding and intent reasoning (Brown et al. 2020; Zhang et al. 2023) by encoding a wide range of human behaviors from their training data. These advancements offer new opportunities for employing LLMs as work assistants, particularly in creating LLM-powered decision support systems (Allen, He, and Gadiraju 2023; Chiang et al. 2024; Wang et al. 2024).

A growing body of research has shown that these LLMs can perform at human levels, or even above, in many experiments (Binz and Schulz 2023b; Dasgupta et al. 2022; Shiffrin and Mitchell 2023) and tasks designed to test different aspects of reasoning (Mahowald et al. 2024). However, empirical findings on their ability to account for ToM are mixed. While some studies show promising results (Strachan et al. 2024), others highlight limitations in accurately reasoning about the mental states of others in different ToM tasks (Ullman 2023). From an empirical standpoint, little is currently understood about whether these models can predict and capture human-like behavioral characteristics, especially human cognitive biases (Mitchell and Krakauer 2023). For instance, an experiment in sequential decision-making that required a trade-off between exploration and exploitation showed that GPT-3 outperformed human subjects by relying heavily on exploitative strategies (Binz and Schulz 2023b). In contrast, people tended to apply a combination of elaborate exploration strategies (Wilson et al. 2014).

Prior research has shown that humans rely on various cognitive mechanisms when making decisions (Gonzalez, Lerch, and Lebiere 2003a; Erev et al. 2010; Lebiere et al. 2013). These cognitive models have been instrumental in understanding the strengths and limitations of human performance and machine learning algorithms (Thomson, Lebiere, and Bennati 2014; Mitsopoulos et al. 2022). With the rise of LLMs, it is unclear how these cognitive models compare to LLMs in predicting human decision-making strategies. Addressing this gap is crucial for gaining deeper insights into LLMs’ potential, providing a cognitive grounding between human users and these models, and guiding the development of LLM systems that can effectively interact with people.

In this work, we investigate the capabilities of LLMs to predict human action strategies in two sequential decision-

making tasks, and compare their performance with a cognitive instance-based learning (IBL) model (Gonzalez, Lerch, and Lebiere 2003a). Based on the theory of experience-based decisions, IBL models simulate human decision-making by incorporating the ACT-R memory mechanisms (Anderson and Lebiere 2014). These models have proven effective in emulating human decisions in various tasks, including gambling choices (Gonzalez and Dutt 2011; Hertwig 2015), complex dynamic resource allocation (Somers, Oltramari, and Lebiere 2020), cybersecurity (Gonzalez et al. 2020), and predicting the actions of other RL agents (Nguyen and Gonzalez 2022).

Our goal is to understand whether LLMs and the cognitive IBL model can predict human action strategies and capture human biases, such as loss aversion, characterized by the tendency to choose sub-optimal goals with fewer step-cost penalties rather than exploring optimal choices. We focus on multi-step, goal-directed decision-making tasks in interactive environments that require balancing exploitative and exploratory actions and handling delayed feedback—essential components of real-life decision processes.

To achieve this, we analyze the discrepancies between the model strategy predictions and real human strategies, which enabled us to uncover the ability of these models to capture the nuances of human behavior in balancing risk and reward during decision-making. We used schema-based and demonstration-based prompts to provide task instructions and users’ action trajectory history from previous trials, allowing pre-trained LLMs to use this in-context information to predict the next action plans in subsequent trials. We employed two open-source LLMs for our experiments: Mistral 7B (Jiang et al. 2023) and Llama-3 70B (the largest of Meta AI’s Llama-3 models with 8B and 70B parameters) (Touvron et al. 2023; Meta 2024). We chose these state-of-the-art open-source LLMs over a closed-source commercial service like GPT-4 (Achiam et al. 2023) as they provide transparency and public access, thus promoting reproducibility and responsible LLM use by giving researchers full access to the network architecture and its pre-trained weights.

Our results show that the lightweight Mistral-7B model outperforms Llama-3 70B and the cognitive IBL model in predicting human strategies. The LLMs quickly improved with more data, though predicting human behavior remained challenging in complex environments. Importantly, we observed that the IBL model better captured human exploratory behavior with limited information and aligned closely with human exploratory strategies under limited information, which reflects the tendency towards risk-averse or “satisficing” behavior (Simon 1956). These findings suggest that integrating LLMs with cognitive architectures could enhance the modeling and understanding of complex human decision-making patterns.

## Related Work

**LLMs for Agent Behavior Modeling.** Generative agents use LLMs to drive their behavior by leveraging the extensive data on human behavior encoded in these models (Brown et al. 2020). Research often relies on templates with few-shot prompts (Gao, Fisch, and Chen 2020) or chain-of-

thought prompts (Wei et al. 2022) to effectively generate behavior based on the agent’s environment. These templates have proven to be effective in control and decision-making tasks. Recent work has shown that LLMs can produce human-like interactions in multi-player games involving natural language communication (Park et al. 2023).

In addition, LLMs have been used to enhance agent modeling with reinforcement learning (RL) agents. Research has shown that integrating feedback into RL models through LLMs provides a learning experience similar to RL with human feedback, without requiring human judgments (Wu et al. 2023b,a; McDonald et al. 2023). LLMs have also improved offline RL, reducing the need for computationally intensive online learning (Shi et al. 2023).

We argue that LLMs offer an opportunity to leverage generative models to understand and predict human behavior. Unlike much existing work that models optimal AI agents, we focus on capturing human behavior.

**LLMs in Theory of Mind Reasoning.** A growing body of research has explored LLMs’ ToM capabilities by testing them with various ToM tasks (Kosinski 2023; Strachan et al. 2024). Results show that leading LLMs can solve 90% of false-belief tasks, sometimes performing at or above human levels, indicating ToM-like abilities. However, Ullman (2023); Shapira et al. (2023) found that LLMs’ performance deteriorates with slight modifications to task structure, highlighting mixed results in this area.

From a modeling perspective, ToM has been used to improve the performance of AI agents in different contexts. Recent studies have applied ToM with LLMs to improve collaboration in multi-agent reinforcement learning (Li et al. 2023; Sun, Huang, and Pompili 2024). We distinguish our work by evaluating the ToM abilities of LLMs to understand human action strategies and biases across various decision-making complexities rather than focusing on learning to infer the intentions of other RL agents.

**Cognitive Modeling and Human Behavior.** Cognitive architectures like ACT-R have demonstrated success in achieving human-level reasoning with limited training instances and in capturing cognitive biases in various decision-making tasks (Anderson and Lebiere 2014; Gonzalez, Lerch, and Lebiere 2003a; Erev et al. 2010; Thomson et al. 2015). Lebiere et al. (2013) showed that the cognitive IBL model predicts whether a person will be risky or risk-averse based on previous trial feedback. Prior research has also indicated that IBL models align with human judgment in predicting RL agents’ goals (Nguyen and Gonzalez 2022) and often serve as a baseline for human behavior (Malloy et al. 2023).

Building on research showing that behavioral traces predict performance (Gadiraju et al. 2019), we explore how cognitive models, particularly the IBL model, compare to LLMs in predicting human action strategies based on previous observations. These processes involve balancing exploitative and exploratory actions with delayed feedback. By comparing the predictive performance of cognitive and LLM-based models, our aim is to establish a baseline for understanding their differences and potential synergies.

## Preliminaries

**Task Scenario.** We studied goal-seeking task environments (Rabinowitz et al. 2018; Nguyen, McDonald, and Gonzalez 2023) that were set in  $10 \times 10$  gridworld mazes containing obstacles and four terminal targets. Each target had a different value, and only one target had the highest value. The reward function for the four terminal objects was drawn randomly from a Dirichlet distribution with a concentration parameter of 0.01. During each episode, the player navigated the grid, making a series of decisions using the common action space (up, down, left, or right) to locate the target with the highest value. The player could consume the targets by moving on top of them. Episodes ended when a target was collected or when the time horizon was reached ( $T_{\max} = 31$ ). The player received points for reaching the target but was also penalized for each movement (-0.01) and for walking into an obstacle (-0.05).

**Task Formulation.** The task is modeled as a partially observable Markov Decision Process (POMDP), represented by the tuple  $\langle \mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{T}, \mathcal{R}, \Omega, \gamma \rangle$ . Here,  $\mathcal{S}$  denotes the state space, with each square in the grid called a state  $s \in \mathcal{S}$ ;  $\mathcal{A}$  is the action space;  $\mathcal{O}$  is the observation space;  $\mathcal{T} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$  is the transition function;  $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$  is the reward function;  $\Omega : \mathcal{S} \rightarrow \mathcal{O}$  is the observation function; and  $\gamma \in [0, 1)$  is the discount factor controlling the player’s emphasis on future rewards compared to immediate rewards.

At every step  $t \in 0, \dots, T_{\max}$ , a player is required to take an action  $a \in \mathcal{A}$  after observing  $o_t \in \mathcal{O}$ . The player receives a reward  $r_t \in \mathbb{R}$  after taking the action as the environment transitions to a new state. Each player follows their policy  $\pi_i$  (i.e., strategy) to decide how to act. By executing its policy  $\pi_i$  in the gridworld  $\mathcal{M}$  following episode  $j$ , the player  $\mathcal{P}_i$  generates a trajectory denoted by  $\mathcal{T}ij = (s_t, a_t)_{t=0}^{T_{\max}}$ .

## Prediction Models

We describe a general approach that adapts LLMs to predict human decisions in sequential goal-directed decision-making tasks. We compare our LLM-based prediction models with the cognitive IBL model. The general framework of our approach is illustrated in Fig. 1a.

### LLM-based Prediction Models

We use LLMs to predict human behaviors in the described task, using both instruction-following and demonstration-based paradigms. Building on prior successes in using LLMs for control settings (Wu et al. 2023a,b; Park et al. 2023), we prompt the LLM to generate a trajectory that human players would take to succeed in the task. Unlike previous research (Wu et al. 2023a; McDonald et al. 2023), which focused on LLMs deciding the provision or value of a reward for optimal performance, we ask the LLM to predict the sequence of actions a human player would take. Specifically, we frame our query to predict the trajectory in the next episode, aiming to match the human strategy.

For each user, we construct a prompt consisting of two main parts: task instructions and sequential interaction histories. The interaction histories include the starting position,

the trajectory taken in previous episodes, and information about any consumed targets and their associated values.

### Prompt Design

#### Instructions:

In a gridworld with obstacles represented by black blocks, a person navigates to find a goal with the highest score among four goals: blue, green, orange, and purple. Movement is restricted to up, down, left, and right directions within the grid. Each episode allows a maximum of 31 steps, with a total of 40 episodes permitted. The score is determined by reaching a target, with a penalty of 0.01 points for each step taken and 0.05 points for colliding with an obstacle. The objective is to locate the highest value target within the grid.

Given the current position at  $(x, y)$ , moving up will result in the new position  $(x, y + 1)$ , moving down will result in the new position  $(x, y - 1)$ , moving right will result in the new position  $(x + 1, y)$ , and moving left will result in the new position  $(x - 1, y)$ .

#### Demonstration:

The  $(x,y)$ -coordinate of the starting position is  $\langle \text{starting\_position} \rangle$ .

The trajectory of episode  $j$ :  $\langle \text{trajectory\_}j \rangle$ .

The player collected goal  $\langle \text{consumed\_goal\_}j \rangle$  with a score of  $\langle \text{consumed\_value\_}j \rangle$ .

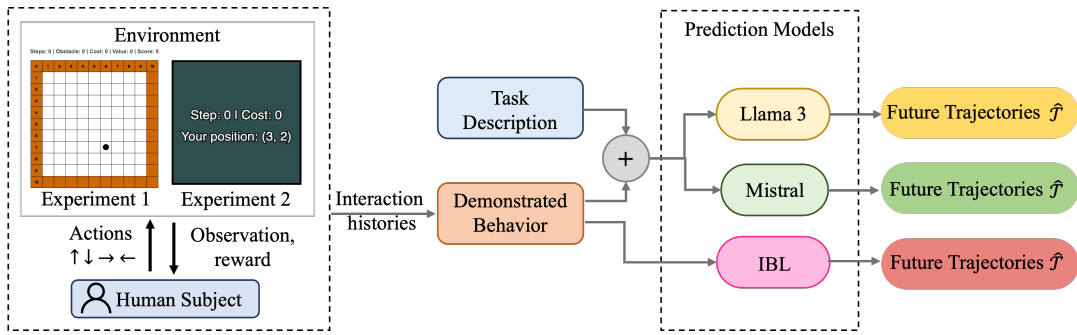
#### Template:

What is the trajectory the player would take in episode  $\langle j+1 \rangle$ ? Please provide only the trajectory in the format of coordinate pairs  $[x,y]$ . Do not explain the reason or include any other words.

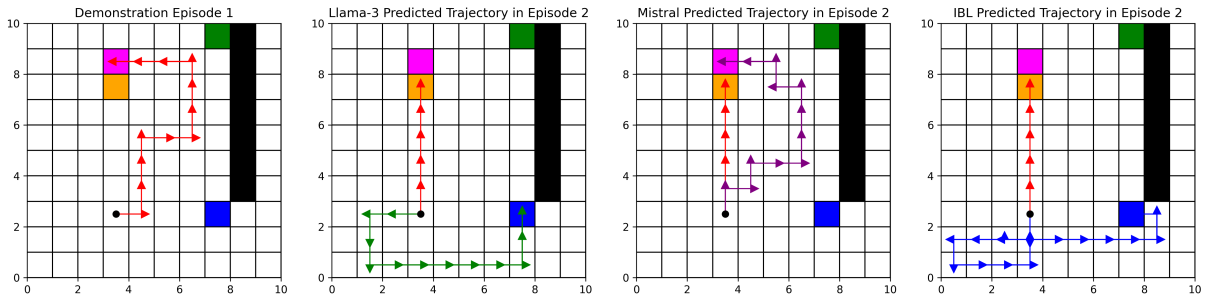
**Output:**  $\langle \text{predicted\_trajectory} \rangle$

In the prompt, we use in-context learning (ICL), a prominent prompting approach for LLMs to solve various tasks (Zhao et al. 2023). This involves incorporating demonstration examples by augmenting the input interaction sequence. Specifically, we pair the prefix of the input interaction sequence with its corresponding successor as examples. For instance, “The trajectory of episode 1:  $[(3, 2), (3, 3), (3, 4), (3, 5)]$ . The player collected goal *blue* with a score of 0.66”. To predict the next actions, we ask, “What is the trajectory the player would take in episode 2?”. This approach enables the LLMs to understand the instructions and output human behavior based on their sequential historical interactions provided.

**POMDP Formulation.** The LLM-based prediction model can be formalized as follows: For each player  $\mathcal{P}_i$ , we aggregate their historical trajectories  $\mathcal{T} = \bigcup_j \mathcal{T}ij = (s_t, a_t)_{t=0}^{T_{\max}}$ ,



(a) The IBL model used the demonstrated behaviors (past trajectories) to predict future behavior, specifically the next-episode trajectory. The LLMs used both the demonstrated behaviors and task descriptions to make such predictions.



(b) An example of the demonstrated trajectory in episode 1 (on the left-most plot), the models' predicted trajectories in the next episode, and the true human player trajectory in that episode (in red). Here, the orange target has the highest value.

Figure 1: An overview of the experiment design.

generated after executing a sequence of actions at each time step  $t$  across  $j$  episodes.

---

#### Algorithm 1: Trajectory Prediction Using LLM

---

- 1: Initialize result storage as a dictionary  $\text{result} = \{\}$
  - 2: **for** each player  $\mathcal{P}_i$  **do**
  - 3: Initialize the context  $\mathcal{C}$  with task instructions and starting position
  - 4: **for** each episode  $j$  **do**
  - 5: Aggregate historical trajectories  $\mathcal{T}_{ij}$  and corresponding consumed targets  $\mathcal{G}_j$  and values  $\mathcal{V}_j$
  - 6: Update the context  $\mathcal{C}$  with  $\mathcal{T}_{ij}$ ,  $\mathcal{G}_j$ , and  $\mathcal{V}_j$
  - 7: Query the LLM model  $M$  with context  $\mathcal{C}$  to predict the next episode's trajectory  $\mathcal{T}_{i(j+1)}$
  - 8: Store the predicted trajectory  $\mathcal{T}_{i(j+1)}$  in  $\text{result}$
  - 9: **end for**
  - 10: **end for**
- 

The context  $\mathcal{C}$  is defined by the combination of task instructions and historical interactions  $\mathcal{T}$ , paired with the consumed targets  $\mathcal{G} = \{g_1, \dots, g_j\}$  corresponding to the trajectory of each episode  $j$ , and the values  $\mathcal{V} = \{v_1, \dots, v_j\}$  associated with obtaining these targets. The LLM model  $M$  uses this context  $\mathcal{C}$  as input to predict the trajectory for the next episode  $\mathcal{T}_{i(j+1)}$ . Essentially, given the context  $\mathcal{C}$  spanning from the first to the  $j$ -th episode, the LLM-based model predicts the trajectory for episode  $j + 1$ . The details of the

algorithm are provided in Algorithm 1.

We note that the output from LLMs may still contain natural language text. We address this by employing text processing methods to parse and ground the generated results in the specified environment. Additionally, we have occasionally observed instances where LLMs produce coordinates that are invalid within the given environment scope. In such cases, we reprocess these illegal outputs to ensure compliance with environmental constraints.

#### Instance-Based Learning (IBL) for Prediction

The IBL model used to compare and predict human behavior is based on Instance-Based Learning Theory (IBLT) (Gonzalez, Lerch, and Lebiere 2003b) for dynamic decision-making, which is connected to the ACT-R cognitive architecture through the activation function, which is used to predict the estimated utility of performing an action in a state based on the utility of similar past experiences held in declarative memory (Thomson et al. 2015).

In IBLT, declarative memory consists of instances  $k = (o, a, x)$  represented by the observation that describes the state of the environment  $o$ , the action performed by the agent  $a$ , and the utility of that action  $x$ . This instance structure can be related to the POMDP environment formulation by taking the state  $s$  as the observation of the agent  $o$ , and the utility  $x$  to be the observed reward  $r$ .

Agent actions are determined by maximizing the value

$V_{k,t}$  of an available action  $a$  in an instance  $k$  performed at time-step  $t$ , calculated using the “blending” function (Gonzalez, Lerch, and Lebiere 2003b):

$$V_{k,t} = \sum_{i=1}^{n_{k,t}} p_{i,k,t} x_{i,k,t} \quad (1)$$

where  $n_{k,t}$  are the previously generated instances held in procedural memory,  $x_{i,k,t}$  are the outcomes of those instances, and  $p_{i,k,t}$  is the probability of retrieving an instance in memory, calculated by Equation 2.

$$p_{i,k,t} = \frac{\exp(\Lambda_{i,k,t}/\tau)}{\sum_{j=1}^{n_{k,t}} \exp(\Lambda_{j,k,t}/\tau)} \quad (2)$$

Further,  $\Lambda_{i,k,t}$  is given by Equation 3.

$$\Lambda_{i,k,t} = \ln \left( \sum_{t' \in T_{i,k,t}} (t - t')^{-d} \right) + \sigma \ln \frac{1 - \xi_{i,k,t}}{\xi_{i,k,t}}, \quad (3)$$

where  $d$  and  $\sigma$  are decay and noise parameters, and  $T_{i,k,t} \subset \{0, \dots, t-1\}$  is the set of previous timesteps where instance  $k$  was stored in memory. The  $\xi_{i,k,t}$  term is used to capture noise in individual differences in memory recall. Due to the relationship between noise  $\sigma$  and temperature  $\tau$  in IBLT, the temperature parameter  $\tau$  is typically set to  $\sigma\sqrt{2}$ . In our experiments, we use all default parameters of  $d = 0.25$  and  $\sigma = 0.5$ . We also set the default utility to 1.0 to encourage exploration through an optimistic prior.

A key aspect of applying IBLT to decision-making is determining the utility of actions. Prior research on temporal credit assignment in IBL models has shown that models assigning equal credit to all decisions closely match human performance (Nguyen, McDonald, and Gonzalez 2023, 2024), which we have consequently chosen to adopt this approach. Formally, if a target is reached in step  $T$ , the target’s value  $R_T$  is assigned to each instance in the trajectory  $\mathcal{T} = (s_t, a_t)_{t=0}^T$ , i.e.,  $x_t = R_T$  for all  $(s_t, a_t)$ . Step-level costs are assigned to each instance if no target is reached.

The IBL prediction model, functioning as an observer, learns by observing past decisions made by human agents. This past experience is incorporated into the model’s memory through pre-populated instances, a mechanism that demonstrates how the IBL model can dynamically represent the development of ToM by observing actions of other learning agents in a gridworld task (Nguyen and Gonzalez 2022). Specifically, for each player  $\mathcal{P}_i$ , the trajectory  $\mathcal{T}_{ij}$  produced by the player, following its policy  $\pi$  in the gridworld  $\mathcal{M}$  after episode  $j$ , is stored in the model memory.

## Methods

Our research aims to determine whether LLMs and the cognitive IBL model can accurately predict human strategic planning in uncertain decision-making environments, formulated as POMDPs, given past interaction histories. These environments require balancing potential high rewards with the risks or losses associated with high-value objects.

We utilized data from two experiments using browser-based interactive gridworld applications that incorporate two

levels of decision complexity. Moreover, we explored how the models predict human strategies under different levels of environment presentation: full grid information in Experiment 1 and restricted grid information in Experiment 2. Our primary research questions are:

- **RQ1:** How accurately can LLMs and the cognitive IBL model predict human action in uncertain decision-making environments based on past interaction histories?
- **RQ2:** To what extent do LLMs and the cognitive IBL model capture human decision biases, such as loss aversion, across different levels of decision complexity?
- **RQ3:** How do different levels of environment presentation (full grid information vs. restricted grid information) affect the accuracy of LLMs and the cognitive IBL model in predicting human decision strategies?

## Experimental Design and Procedure

The two experiments used the same gridworlds, but the information provided to participants varied. Human subjects were presented with gridworlds randomly chosen from a set of 100 grids, with the selection based on the decision complexity level assigned to each participant’s condition.

Participants were recruited from Amazon Mechanical Turk and provided informed consent before completing each session. After receiving instructions, participants completed 40 episodes in the same gridworld environment, with each session lasting 15-30 minutes. They received a base payment of \$1.50 and could earn up to \$3.00 in bonuses based on their accumulated scores. The studies, approved by our institution’s IRB, employed a between-subjects design and were preregistered with the Open Science Framework for Experiment 1<sup>1</sup> and Experiment 2<sup>2</sup>.

**Decision Complexity.** The experiments manipulate the level of decision complexity defined by the trade-offs between the highest-value target and the nearest distractor relative to the agent’s initial spawn location in the gridworld (Nguyen and Gonzalez 2020). This complexity is quantified by  $\Delta_d = d - d'$ , where  $d$  is the distance to the highest value target, and  $d'$  is the distance to the nearest distractor. Higher  $\Delta_d$  values represent greater complexity, posing a strategic dilemma for agents: pursue a distant high-reward target or opt for a closer, less valuable one. Fig. 2 illustrates simple and complex decision scenarios.

**Experiment 1: Full Grid Information.** Participants viewed the full grid interface (Fig. 1a), with their current position indicated by a black dot. After each move, the new location content (empty cell, obstacle, or target) was revealed.

*Participants.* A total of 206 participants: 102 in the Simple condition (age:  $36.5 \pm 10.3$ ; 34 female) and 104 in the Complex condition (age:  $37.9 \pm 10.8$ ; 37 female).

**Experiment 2: Restricted Grid Information.** Participants received limited information, viewing only one cell at a time (Fig. 1a). They were informed of their current  $(x, y)$  position, the steps taken, and the immediate cost or reward

<sup>1</sup>Experiment 1: <https://osf.io/2ycm6>

<sup>2</sup>Experiment 2: <https://osf.io/hxfyq>

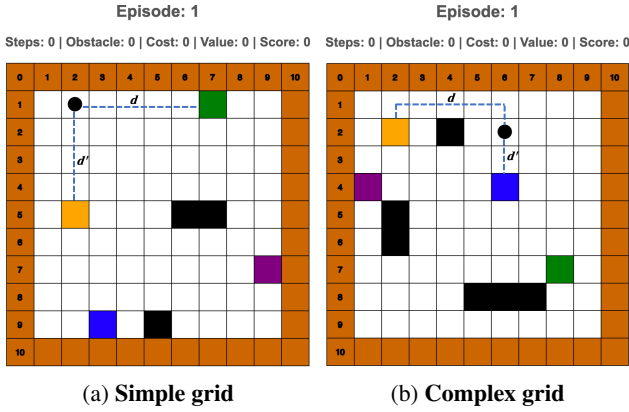


Figure 2: Example grids for simple and complex conditions:(a) “green” is the highest value target with “orange” as the distractor ( $\Delta_d = 1$ ); (b) “orange” is the target, “blue” is the distractor ( $\Delta_d = 4$ ).

of the previous step. All other information about the shape and size of the grid were concealed.

*Participants.* A total of 194 participants: 99 in the Simple condition (age:  $37.7 \pm 11.8$ ; 40 female) and 95 in the Complex condition (age:  $38.2 \pm 11.3$ ; 30 female).

**Model Implementation.** We implemented the cognitive IBL model using the SpeedyIBL library (Nguyen, Phan, and Gonzalez 2023b) with default parameters. The LLM-based models were accessed through the Ollama REST API. See: <https://github.com/ollama/ollama>.

## Objective Measures

**Trajectory Divergence:** We used Kullback-Leibler (KL) divergence to measure the difference between the trajectory distribution of human subjects and that predicted by the model. Each trajectory is converted into a probability distribution over the grid cells, normalized to sum to 1. The KL divergence from  $Q$  (predicted trajectory) to  $P$  (human trajectory) is defined as:

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \quad (4)$$

where  $i$  indexes each possible state in the trajectory grid. Low KL divergence indicates that the model closely matches human behavior.

**Prediction Accuracy:** This measures the percentage of episodes where the predicted target, derived from the last coordinate of the predicted trajectory, matches the target consumed by human players.

**Exploration Entropy Difference:** This metric measures the difference in entropy of the distribution over how often each target is explored by humans (human goal entropy) and by the model (predicted goal entropy) in the first 10 episodes. The entropy difference is determined by subtracting the human goal entropy from the predicted goal entropy. Lower entropy difference suggests that the model’s exploration behavior closely aligns with human behavior, indicating similar patterns in exploring targets.

## Analysis

We compared the LLM-based models Llama-3 70B (Meta 2024) and Mistral 7B (Jiang et al. 2023) with the cognitive IBL model across a  $2 \times 2$  study design in two experiments, each with simple and complex decision complexities.

**Trajectory Divergence.** Table 1 shows that in both experiments, the Mistral model with 7B consistently achieved the lowest KL divergence in simple and complex conditions, indicating better alignment with human trajectories compared to Llama-3 with 70B and the cognitive IBL model. When comparing simple and complex decision settings, predicting human strategies is more challenging in complex environments, as evidenced by the increased KL divergence.

When comparing Experiments 1 and 2, we observe that both Llama-3 and Mistral show better alignment with human trajectories under restricted information, as indicated by the decreased KL divergence in Experiment 2, while this is not the case for the IBL model. This improved alignment for the LLMs can be attributed to their pre-training on vast datasets and the incorporation of instructions, which enhance their contextual understanding. When human participants receive restricted grid information, as in Experiment 2, they may adopt more predictable strategies that LLM models can more easily capture. In contrast, the IBL model shows poorer alignment with human trajectories under restricted information as it struggles to distinguish the differences that human players encountered in the two different conditions.

Experiment	Condition	Model	Mean KL $\pm$ SE
Experiment 1	Simple	Llama-3	$6.189 \pm 0.128$
		<b>Mistral</b>	<b><math>4.895 \pm 0.117</math></b>
		IBL	$5.244 \pm 0.122$
	Complex	Llama-3	$8.385 \pm 0.126$
		<b>Mistral</b>	<b><math>6.524 \pm 0.118</math></b>
		IBL	$7.216 \pm 0.119$
Experiment 2	Simple	Llama-3	$5.901 \pm 0.127$
		<b>Mistral</b>	<b><math>4.641 \pm 0.114</math></b>
		IBL	$6.061 \pm 0.128$
	Complex	Llama-3	$7.416 \pm 0.129$
		<b>Mistral</b>	<b><math>5.910 \pm 0.120</math></b>
		IBL	$8.853 \pm 0.123$

Table 1: Average KL Divergence for each model under different conditions in Experiment 1 and Experiment 2.

Fig. 3 shows the average KL divergence per episode. As expected, in both experiments and conditions, all three models show a decreasing KL divergence trend, indicating better alignment with human trajectories as episodes and demonstrated samples increase. Notably, in the simple condition, the IBL model performs comparably to Mistral and even outperforms Llama-3 in the first 10 episodes, suggesting effective capturing of human actions with few-shot examples.

**Prediction Accuracy.** The results of Table 2 indicate that the cognitive IBL model is the most accurate in predicting human target consumption under simple conditions in Experiment 1, achieving the highest prediction accuracy ( $0.688 \pm 0.007$ ) compared to Llama-3 and Mistral.

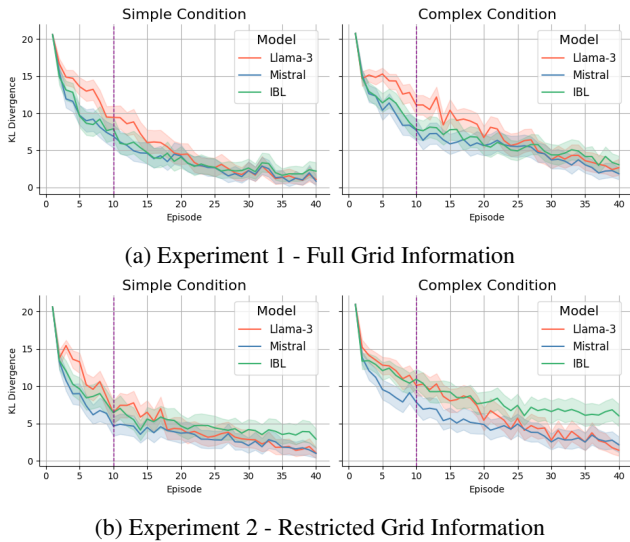


Figure 3: Average KL divergence per episode for all models in both conditions of Experiments 1 and 2, with shaded areas indicating standard error at 95% confidence intervals. Lower KL divergence suggests better alignment.

In both experiments, all models exhibit lower prediction accuracy in complex settings compared to simple settings. In Experiment 2, while Llama-3 and Mistral maintain similar accuracies to those in Experiment 1, the IBL model shows a decrease in prediction accuracy, especially in complex conditions. This suggests that the IBL model may struggle to capture the high variability of human decisions when participants are provided with limited information.

Experiment	Condition	Model	Accuracy $\pm$ SE
Experiment 1	Simple	Llama-3	$0.651 \pm 0.007$
		Mistral	$0.658 \pm 0.007$
		<b>IBL</b>	<b><math>0.688 \pm 0.007</math></b>
	Complex	Llama-3	$0.523 \pm 0.008$
		<b>Mistral</b>	<b><math>0.594 \pm 0.008</math></b>
		IBL	$0.560 \pm 0.008$
Experiment 2	Simple	Llama-3	$0.652 \pm 0.008$
		<b>Mistral</b>	<b><math>0.655 \pm 0.008</math></b>
		IBL	$0.608 \pm 0.008$
	Complex	Llama-3	$0.542 \pm 0.008$
		<b>Mistral</b>	<b><math>0.602 \pm 0.008</math></b>
		IBL	$0.382 \pm 0.008$

Table 2: Average Prediction Accuracy for each model under different conditions in Experiment 1 and Experiment 2.

Notably, in the first 10 episodes, the IBL model shows high prediction accuracy, surpassing both Llama-3 and Mistral in both simple and complex conditions, as shown in Fig. 4. This suggests that the IBL model effectively captures the initial exploration strategies of human players, learning quickly and adapting well with a few samples. However, as more episodes are added, the LLMs increase and quickly improve their prediction accuracy, eventually matching or exceeding the performance of the IBL model.

This finding indicates that while the IBL model aligns closely with human decisions during initial exploration, LLMs can leverage larger amounts of data to refine their predictions and better capture human actions over time.

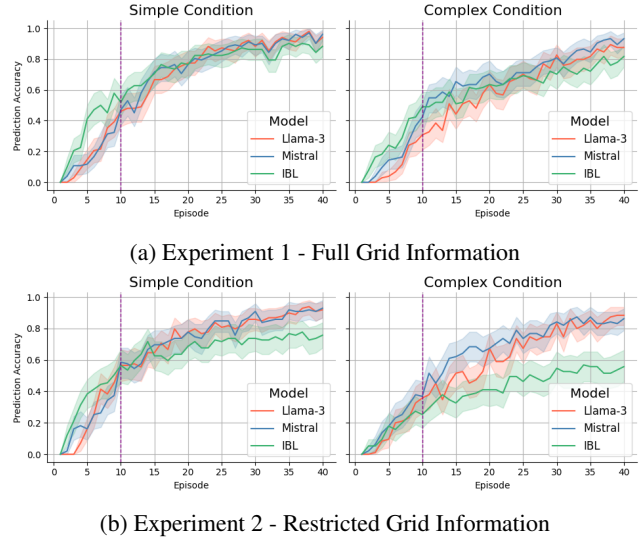


Figure 4: Average prediction accuracy per episode for all models in both conditions of Experiments 1 and 2. Shaded areas indicate the standard error. Higher prediction accuracy suggests better alignment with human target consumption.

**Exploration Entropy Difference.** Fig. 5 shows the differences in the early exploratory behavior between each model and humans within the first 10 episodes. In Experiment 1, IBL shows a negative entropy difference in both simple and complex conditions, suggesting that it explores less than humans. On the other hand, Llama-3 and Mistral exhibit entropy differences close to zero or slightly positive, with Mistral aligning most closely with human exploration behavior.

In Experiment 2, IBL again shows negative or near-zero entropy differences, particularly under simple conditions, indicating conservative exploration. This aligns with human behavior, suggesting that with restricted information, people adopt risk-averse strategies well captured by the IBL model.

The results indicate that the IBL model, with default parameters, naturally exhibits a conservative exploration strategy, aligning with prior findings regarding its risk-aversion tendency compared to winner-take-all instances, which is the optimal strategy overall (Lebiere et al. 2013). This characteristic effectively captures human exploration behavior when information is restricted. In such settings, people are limited in their exploration and often settle on the closest sub-optimal target rather than seeking the highest-value target, reflecting a human bias towards risk aversion. In contrast, Mistral consistently demonstrates exploratory patterns that are more aligned with those of humans.

## Discussions

Our results show that Mistral-7B outperforms both Llama-3 70B and the cognitive IBL model in predicting human

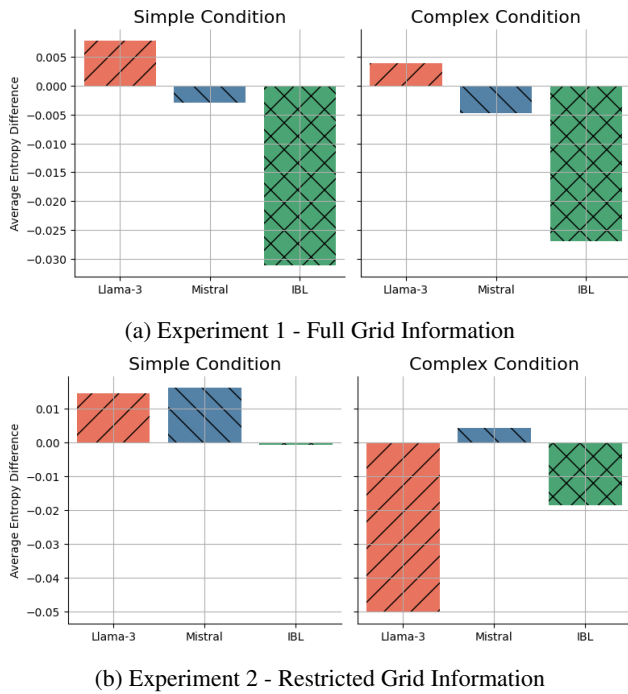


Figure 5: Average entropy difference for all models in both conditions of Experiment 1 and 2. Positive values indicate more exploration than humans; negative values indicate less and near-zero values show alignment with human behavior.

strategies. We also found that predicting human behavior becomes more challenging in complex decision environments with high tension between costs and rewards. Moreover, the cognitive IBL model effectively captures initial human exploratory behavior with minimal demonstration samples. However, as more samples are provided, LLMs quickly improve their prediction accuracy by leveraging their extensive pre-training on vast datasets and instructions that enhance their contextual understanding. Importantly, our findings indicate that the IBL model, with its inherent risk aversion tendency, is closely aligned with human exploration strategies under limited information conditions. It effectively captures the human tendency toward risk-aversion, “satisficing” behavior (Simon 1956), where people often choose an option that is satisfactory rather than optimal.

### Implications for Trust and Synergy in AI-assisted Systems with LLMs and Cognitive Models

One of our findings is that the lightweight LLM Mistral-7B outperforms the widely recognized Llama-3 70B and the cognitive IBL model in predicting human action strategies in various decision-making settings. This highlights the potential of using lightweight open source LLMs to develop reliable AI systems. Although much research on human interaction powered by LLM focuses on black-box models such as ChatGPT, our study emphasizes the capabilities of open-source pre-trained LLMs, which allow fine-tuning and provide access to network architecture and weights, facilitating

cognitive-plausible understanding and integration (Binz and Schulz 2023a; Malloy and Gonzalez 2024). Previous work shows that humans initially underestimate AI and then over-trust it with more experience, highlighting the need for explicit calibration of AI competencies (Rechkemmer and Yin 2022; Bućinca, Malaya, and Gajos 2021). Leveraging open-source models can create trustworthy, accurate, explainable, and aligned LLM-powered systems, enhancing their acceptance and effectiveness in real-world applications.

Our experiments with POMDPs, focused on simplicity and ease of control, have the potential to be generalized to richer domains where decisions involve uncertainty and incomplete information. This approach can be applied to applications such as robotics, autonomous systems, and financial decision-making, where adaptive strategies are necessary to handle unpredictability and dynamic environments.

In the context of AI-assisted human decision-making, our findings highlight the utility of the cognitive IBL model in capturing initial human exploratory behavior and the tendency toward loss aversion in high cost-reward tension scenarios without requiring large amounts of training data. Conversely, LLMs can quickly learn and predict human strategies as more data becomes available, indicating a synergy between these models. For instance, LLMs can support cognitive models by synthesizing large amounts of information and serving as knowledge repositories to construct representations of the environment (Wu et al. 2023a; Binz and Schulz 2023a). Cognitive IBL models, which can predict human strategies with few-shot learning, have proven beneficial when integrated with multi-agent deep reinforcement learning techniques to enhance coordination in multi-agent systems with stochastic rewards (Nguyen, Phan, and Gonzalez 2023a). Thus, cognitive models can help LLMs adapt early on and further personalize their responses to human users (Malloy and Gonzalez 2024; Thomson and Bastian 2023). This synergy would benefit effective human-AI teaming, where AI evolves alongside human learning and adaptation to support human decision-making.

### Limitations and Future Work

Our experiments were simple and aimed to shed light on the capabilities of open-source LLMs and cognitive architectures in predicting human behavior. There is considerable room for further investigation. First, our study focused on purely open-source LLMs, so caution should be exercised when extrapolating our findings to closed-source commercial services like ChatGPT, which may exhibit different performance levels. Second, we used the vanilla versions of these models without fine-tuning them to human data, particularly in cognitive IBL models, where an equal credit assignment mechanism leads to conservative exploratory strategies. Future research could enhance these models to better match human behavior and uncover new behavioral structures. Finally, while our study highlights the strengths of different models, it does not fully explore the integration of LLMs and cognitive models in a cohesive framework, which we consider as future work. Future research would also benefit from investigating the potential of these models in interactive systems to predict human decisions in real time.

## Acknowledgements

This research was supported by the University of Dayton Research Council Seed Grant and the UD/UDRI Summer Research Fellows Program. Cleotilde Gonzalez was supported by the AI Research Institutes Program funded by the National Science Foundation under AI Institute for Societal Decision Making (AI-SDM), Award No. 2229881.

## References

- Abbeel, P.; and Ng, A. Y. 2004. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on Machine learning*, 1.
- Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F. L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- Allen, G.; He, G.; and Gadiraju, U. 2023. Power-up! What Can Generative Models Do for Human Computation Workflows? *arXiv preprint arXiv:2307.02243*.
- Anderson, J. R.; and Lebiere, C. J. 2014. *The atomic components of thought*. Psychology Press.
- Baker, C. L.; Jara-Ettinger, J.; Saxe, R.; and Tenenbaum, J. B. 2017. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, 1(4): 0064.
- Bansal, G.; Nushi, B.; Kamar, E.; Weld, D. S.; Lasecki, W. S.; and Horvitz, E. 2019. Updates in human-ai teams: Understanding and addressing the performance/compatibility tradeoff. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 2429–2437.
- Binz, M.; and Schulz, E. 2023a. Turning large language models into cognitive models. *arXiv preprint arXiv:2306.03917*.
- Binz, M.; and Schulz, E. 2023b. Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, 120(6): e2218523120.
- Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901.
- Buçinca, Z.; Malaya, M. B.; and Gajos, K. Z. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1): 1–21.
- Chiang, C.-W.; Lu, Z.; Li, Z.; and Yin, M. 2024. Enhancing AI-Assisted Group Decision Making through LLM-Powered Devil’s Advocate. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*, 103–119.
- Chowdhery, A.; Narang, S.; Devlin, J.; Bosma, M.; Mishra, G.; Roberts, A.; Barham, P.; Chung, H. W.; Sutton, C.; Gehrmann, S.; et al. 2023. Palm: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240): 1–113.
- Dasgupta, I.; Lampinen, A. K.; Chan, S. C.; Creswell, A.; Kumaran, D.; McClelland, J. L.; and Hill, F. 2022. Language models show human-like content effects on reasoning. *arXiv preprint arXiv:2207.07051*.
- Erev, I.; Ert, E.; Roth, A. E.; Haruvy, E.; Herzog, S. M.; Hau, R.; Hertwig, R.; Stewart, T.; West, R.; and Lebiere, C. 2010. A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, 23(1): 15–47.
- Gadiraju, U.; Demartini, G.; Kawase, R.; and Dietze, S. 2019. Crowd anatomy beyond the good and bad: Behavioral traces for crowd worker modeling and pre-selection. *Computer Supported Cooperative Work (CSCW)*, 28: 815–841.
- Gao, T.; Fisch, A.; and Chen, D. 2020. Making pre-trained language models better few-shot learners. *arXiv preprint arXiv:2012.15723*.
- Gonzalez, C.; Aggarwal, P.; Cranford, E. A.; and Lebiere, C. 2020. Design of dynamic and personalized deception: A research framework and new insights for cyberdefense. In *Proceedings of the 53rd hawaii international conference on system sciences*, volume 1834.
- Gonzalez, C.; and Dutt, V. 2011. Instance-based learning: Integrating decisions from experience in sampling and repeated choice paradigms. *Psychological Review*, 118(4): 523–51.
- Gonzalez, C.; Lerch, J. F.; and Lebiere, C. 2003a. Instance-based learning in dynamic decision making. *Cognitive Science*, 27(4): 591–635.
- Gonzalez, C.; Lerch, J. F.; and Lebiere, C. 2003b. Instance-based learning in dynamic decision making. *Cognitive Science*, 27(4): 591–635.
- Hertwig, R. 2015. Decisions from experience. *The Wiley Blackwell handbook of judgment and decision making*, 2: 239–267.
- Hoffman, G.; Bhattacharjee, T.; and Nikolaidis, S. 2023. Inferring Human Intent and Predicting Human Action in Human–Robot Collaboration. *Annual Review of Control, Robotics, and Autonomous Systems*, 7.
- Jiang, A. Q.; Sablayrolles, A.; Mensch, A.; Bamford, C.; Chaplot, D. S.; Casas, D. d. I.; Bressand, F.; Lengyel, G.; Lample, G.; Saulnier, L.; et al. 2023. Mistral 7B. *arXiv preprint arXiv:2310.06825*.
- Kosinski, M. 2023. Theory of mind may have spontaneously emerged in large language models. *arXiv preprint arXiv:2302.02083*, 4: 169.
- Lebiere, C.; Pirolli, P.; Thomson, R.; Paik, J.; Rutledge-Taylor, M.; Staszewski, J.; and Anderson, J. R. 2013. A functional model of sensemaking in a neurocognitive architecture. *Computational intelligence and neuroscience*, 2013(1): 921695.
- Li, H.; Chong, Y. Q.; Stepputtis, S.; Campbell, J.; Hughes, D.; Lewis, M.; and Sycara, K. 2023. Theory of mind for multi-agent collaboration via large language models. *arXiv preprint arXiv:2310.10701*.
- Mahowald, K.; Ivanova, A. A.; Blank, I. A.; Kanwisher, N.; Tenenbaum, J. B.; and Fedorenko, E. 2024. Dissociating

- language and thought in large language models. *Trends in Cognitive Sciences*.
- Malloy, T.; Du, Y.; Fang, F.; and Gonzalez, C. 2023. Accounting for Transfer of Learning Using Human Behavior Models. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 11, 115–126.
- Malloy, T.; and Gonzalez, C. 2024. Applying Generative Artificial Intelligence to cognitive models of decision making. *Frontiers in Psychology*, 15: 1387948.
- McDonald, C.; Malloy, T.; Nguyen, T. N.; and Gonzalez, C. 2023. Exploring the Path from Instructions to Rewards with Large Language Models in Instance-Based Learning. In *Proceedings of the AAAI Symposium Series*, volume 2, 334–339.
- Meta. 2024. Meta LLaMA-3. Available online.
- Mitchell, M.; and Krakauer, D. C. 2023. The debate over understanding in AI’s large language models. *Proceedings of the National Academy of Sciences*, 120(13): e2215907120.
- Mitsopoulos, K.; Somers, S.; Schooler, J.; Lebiere, C.; Pirolli, P.; and Thomson, R. 2022. Toward a psychology of deep reinforcement learning agents using a cognitive architecture. *Topics in Cognitive Science*, 14(4): 756–779.
- Nguyen, T. N.; and Gonzalez, C. 2020. Effects of decision complexity in goal-seeking gridworlds: A comparison of instance-based learning and reinforcement learning agents. In *Proceedings of the 18th intl. conf. on cognitive modelling*.
- Nguyen, T. N.; and Gonzalez, C. 2022. Theory of mind from observation in cognitive models and humans. *Topics in cognitive science*, 14(4): 665–686.
- Nguyen, T. N.; McDonald, C.; and Gonzalez, C. 2023. Credit assignment: Challenges and opportunities in developing human-like ai agents. *arXiv preprint arXiv:2307.08171*.
- Nguyen, T. N.; McDonald, C.; and Gonzalez, C. 2024. Temporal Credit Assignment in Instance-Based Learning Models. *Proceedings of the AAAI 2024 Spring Symposium Series*, 3(1): 54–57.
- Nguyen, T. N.; Phan, D. N.; and Gonzalez, C. 2023a. Learning in Cooperative Multiagent Systems Using Cognitive and Machine Models. *ACM Transactions on Autonomous and Adaptive Systems*, 18(4): 1–22.
- Nguyen, T. N.; Phan, D. N.; and Gonzalez, C. 2023b. SpeedyIBL: A comprehensive, precise, and fast implementation of instance-based learning theory. *Behavior Research Methods*, 55(4): 1734–1757.
- Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35: 27730–27744.
- Park, J. S.; O’Brien, J. C.; Cai, C. J.; Morris, M. R.; Liang, P.; and Bernstein, M. S. 2023. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442*.
- Rabinowitz, N.; Perbet, F.; Song, F.; Zhang, C.; Eslami, S. A.; and Botvinick, M. 2018. Machine theory of mind. In *International conference on machine learning*, 4218–4227. PMLR.
- Rastogi, C.; Leqi, L.; Holstein, K.; and Heidari, H. 2023. A Taxonomy of Human and ML Strengths in Decision-Making to Investigate Human-ML Complementarity. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 11, 127–139.
- Rechkemmer, A.; and Yin, M. 2022. When confidence meets accuracy: Exploring the effects of multiple performance indicators on trust in machine learning models. In *Proceedings of the 2022 chi conference on human factors in computing systems*, 1–14.
- Shapira, N.; Levy, M.; Alavi, S. H.; Zhou, X.; Choi, Y.; Goldberg, Y.; Sap, M.; and Shwartz, V. 2023. Clever hans or neural theory of mind? stress testing social reasoning in large language models. *arXiv preprint arXiv:2305.14763*.
- Shi, R.; Liu, Y.; Ze, Y.; Du, S. S.; and Xu, H. 2023. Unleashing the Power of Pre-trained Language Models for Offline Reinforcement Learning. *arXiv preprint arXiv:2310.20587*.
- Shiffrin, R.; and Mitchell, M. 2023. Probing the psychology of AI models. *Proceedings of the National Academy of Sciences*, 120(10): e2300963120.
- Simon, H. A. 1956. Rational choice and the structure of the environment. *Psychological review*, 63(2): 129.
- Somers, S.; Oltramari, A.; and Lebiere, C. 2020. Cognitive Twin: A Cognitive Approach to Personalized Assistants. In *Aaai spring symposium: Combining machine learning with knowledge engineering (1)*.
- Strachan, J. W.; Albergo, D.; Borghini, G.; Pansardi, O.; Scaliti, E.; Gupta, S.; Saxena, K.; Rufo, A.; Panzeri, S.; Manzi, G.; et al. 2024. Testing theory of mind in large language models and humans. *Nature Human Behaviour*, 1–11.
- Sun, C.; Huang, S.; and Pompili, D. 2024. LLM-based Multi-Agent Reinforcement Learning: Current and Future Directions. *arXiv preprint arXiv:2405.11106*.
- Thomson, R.; Lebiere, C.; Anderson, J. R.; and Staszewski, J. 2015. A general instance-based learning framework for studying intuitive decision-making in a cognitive architecture. *Journal of Applied Research in Memory and Cognition*, 4(3): 180–190.
- Thomson, R.; Lebiere, C.; and Bennati, S. 2014. Human, model and machine: a complementary approach to big data. In *Proceedings of the 2014 Workshop on Human Centered Big Data Research*, 27–31.
- Thomson, R. H.; and Bastian, N. D. 2023. Integrating Cognitive Architectures with Foundation Models: Cognitively-Guided Few-Shot Learning to Support Trusted Artificial Intelligence. In *Proceedings of the AAAI Symposium Series*, volume 2, 409–414.
- Torabi, F.; Warnell, G.; and Stone, P. 2018. Behavioral cloning from observation. *arXiv preprint arXiv:1805.01954*.
- Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Ullman, T. 2023. Large language models fail on trivial alterations to theory-of-mind tasks. *arXiv preprint arXiv:2302.08399*.

Wang, B.; Li, Y.; Lv, Z.; Xia, H.; Xu, Y.; and Sodhi, R. 2024. LAVE: LLM-Powered Agent Assistance and Language Augmentation for Video Editing. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*, 699–714.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35: 24824–24837.

Wilson, R. C.; Geana, A.; White, J. M.; Ludvig, E. A.; and Cohen, J. D. 2014. Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of experimental psychology: General*, 143(6): 2074.

Wu, Y.; Fan, Y.; Liang, P. P.; Azaria, A.; Li, Y.; and Mitchell, T. M. 2023a. Read and reap the rewards: Learning to play atari with the help of instruction manuals. *arXiv preprint arXiv:2302.04449*.

Wu, Y.; Min, S. Y.; Prabhunoye, S.; Bisk, Y.; Salakhutdinov, R.; Azaria, A.; Mitchell, T.; and Li, Y. 2023b. SPRING: GPT-4 Out-performs RL Algorithms by Studying Papers and Reasoning. *arXiv preprint arXiv:2305.15486*.

Zhang, J.; Xie, R.; Hou, Y.; Zhao, W. X.; Lin, L.; and Wen, J.-R. 2023. Recommendation as instruction following: A large language model empowered recommendation approach. *arXiv preprint arXiv:2305.07001*.

Zhang, R.; McNeese, N. J.; Freeman, G.; and Musick, G. 2021. "An ideal human" expectations of AI teammates in human-AI teaming. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3): 1–25.

Zhao, W. X.; Zhou, K.; Li, J.; Tang, T.; Wang, X.; Hou, Y.; Min, Y.; Zhang, B.; Zhang, J.; Dong, Z.; et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223*.