

# MultiMind: Enhancing Werewolf Agents with Multimodal Reasoning and Theory of Mind

Zheng Zhang\*

The Hong Kong University of Science  
and Technology (Guangzhou)  
Guangzhou, China  
zzhang302@connect.hkust-gz.edu.cn

Nuoqian Xiao\*†

The Hong Kong University of Science  
and Technology (Guangzhou)  
Guangzhou, China  
xiaoqian@sjtu.edu.cn

Qi Chai

The Hong Kong University of Science  
and Technology (Guangzhou)  
Guangzhou, China  
ericedu@stu.xjtu.edu.cn

Deheng Ye‡

Tencent  
Shenzhen, China  
dericye@tencent.com

Hao Wang‡

The Hong Kong University of Science  
and Technology (Guangzhou)  
Guangzhou, China  
haowang@hkust-gz.edu.cn

## Abstract

Large Language Model (LLM) agents have demonstrated impressive capabilities in social deduction games (SDGs) like Werewolf, where strategic reasoning and social deception are essential. However, current approaches remain limited to textual information, ignoring crucial multimodal cues such as facial expressions and tone of voice that humans naturally use to communicate. Moreover, existing SDG agents primarily focus on inferring other players' identities without modeling how others perceive themselves or fellow players. To address these limitations, we use One Night Ultimate Werewolf (ONUW) as a testbed and present MultiMind, the first framework integrating multimodal information into SDG agents. MultiMind processes facial expressions and vocal tones alongside verbal content, while employing a Theory of Mind (ToM) model to represent each player's suspicion levels toward others. By combining this ToM model with Monte Carlo Tree Search (MCTS), our agent identifies communication strategies that minimize suspicion directed at itself. Through comprehensive evaluation in both agent-versus-agent simulations and studies with human players, we demonstrate MultiMind's superior performance in gameplay. Our work presents a significant advancement toward LLM agents capable of human-like social reasoning across multimodal domains. Our code is available at <https://github.com/CjangCjengh/onuw>.

## CCS Concepts

- Computing methodologies → Natural language generation; Activity recognition and understanding; Speech recognition.

\*Equal contribution.

†This work was done during internship at HKUST(GZ).

‡Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MM '25, Dublin, Ireland.

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-2035-2/2025/10

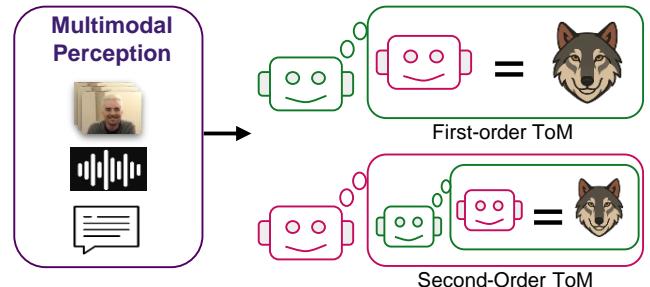
<https://doi.org/10.1145/3746027.3755752>

## Keywords

Social Deduction Games, Large Language Models, Multimodal Reasoning, Emotion Recognition

## ACM Reference Format:

Zheng Zhang, Nuoqian Xiao, Qi Chai, Deheng Ye, and Hao Wang. 2025. MultiMind: Enhancing Werewolf Agents with Multimodal Reasoning and Theory of Mind. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3746027.3755752>



**Figure 1: An example of Theory of Mind (ToM).** First-order ToM involves a player inferring other players' identities, while second-order ToM extends to reasoning about the individual opinions of others.

## 1 Introduction

Large Language Model (LLM) agents have made significant progress in simulating human-like decision-making and social behavior across various domains. Recent research has demonstrated their ability to play games such as StarCraft [20, 28] and MineCraft [17, 23, 31]. Within this broad research landscape, a promising direction has developed focusing on games that require not just strategic thinking but also sophisticated social reasoning. Social deduction games (SDGs) such as Werewolf [36–38, 40], Avalon [14, 19, 29, 33], and Jubensha [18, 35] have proven to be challenging testbeds for LLM agent research. These games demand nuanced reasoning and

strategic communication based on hidden information, which are skills that more closely resemble human social intelligence.

Despite notable progress in developing LLM agents for SDGs, current approaches exhibit significant limitations. Most existing agents are constrained to processing purely textual information, overlooking crucial multimodal cues such as facial expressions and tone of voice. In real-world settings, these non-verbal signals often reveal underlying intentions, emotional states, and potential deception that complement verbal exchanges. While some research has begun incorporating multimodal data in SDG analysis [13, 16], these approaches primarily focus on retrospective analysis rather than active gameplay participation.

Additionally, current methods typically attempt to infer other players' identities, but fail to model how others perceive the identity of the agent itself or of other players. As shown in Figure 1, this layered reasoning structure, known as Theory of Mind (ToM), is the cognitive ability to understand and identify mental states in oneself and others. The lack of systematic ToM modeling capabilities prevents agents from engaging in the multi-level strategic reasoning that human players naturally employ in complex interactions.

In this paper, we focus on One Night Ultimate Werewolf (ONUW) following recent research [12, 13], a variant of the traditional Werewolf game. ONUW is particularly suitable for our research as it has available multimodal datasets [13], such that we may use the multimodal cues to conduct more comprehensive reasoning. Besides, it concentrates gameplay into a single night phase followed by one day phase for discussion and voting, placing emphasis on strategic communication and social reasoning. This creates an ideal testbed for exploring multimodal reasoning in social settings where verbal and non-verbal cues become crucial for communication.

To this end, we present MultiMind, a novel framework that enhances LLM agents for ONUW by integrating multimodal information and ToM reasoning. MultiMind converts facial expressions and tone of voice into textual descriptions that capture emotional signals, enabling our agent to process multimodal information. We further implement ToM reasoning by encoding player statements into discrete action representations, which predicts the belief distributions that represent each player's suspicion levels toward others. Then, we employ Monte Carlo Tree Search (MCTS) to optimize communication strategies based on this ToM model, identifying utterances that minimize suspicion directed at the agent.

The training of the ToM model follows a two-stage approach. We initially use synthetic data from LLM-based agent self-play, then fine-tune on human gameplay data. Through comprehensive evaluation in both agent simulations and studies with human players, we demonstrate substantial improvements in gameplay performance while generating more convincing and strategically sound communication patterns. MultiMind presents a notable step toward LLM agents that can effectively engage in complex social reasoning across multimodal domains.

Our main contributions can be summarized as follows:

- To our best knowledge, we are the first to develop a framework that integrates multimodal information into SDG agents, enhancing their realism and strategic capabilities.
- We propose a novel approach combining ToM modeling with MCTS to optimize communication strategies, enabling

agents to reason about belief states and strategically minimize suspicion directed at themselves.

- We demonstrate the effectiveness of our approach through comprehensive evaluation in both agent-versus-agent simulations and studies with human players, showing improvements in gameplay performance.

## 2 Related Work

### 2.1 Social Deduction Game Agent

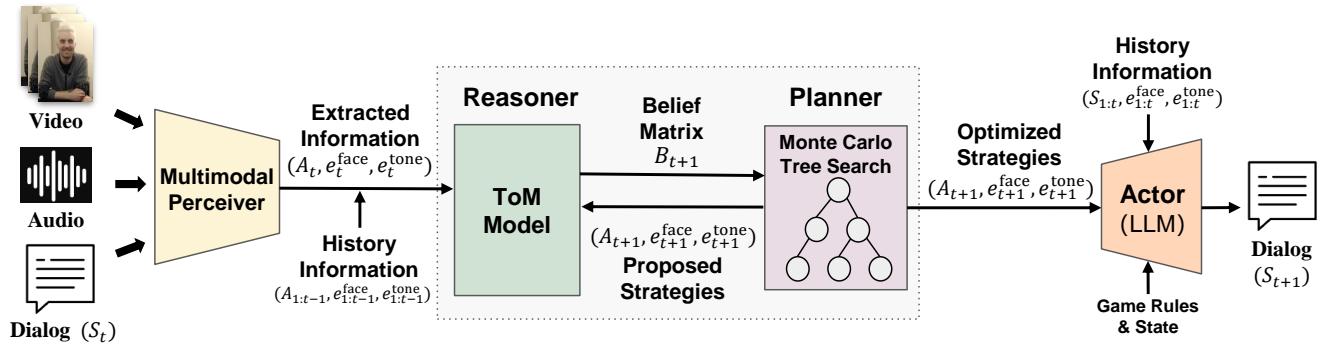
Social deduction games (SDGs) have emerged as valuable testbeds for AI research. Early research [11, 21, 27, 34] on SDG agents employs rule-based or learning-based methods for decision-making. These agents use predefined protocols for communication rather than natural language. With the emergence of LLMs, more sophisticated agents have been developed across various SDGs. For Avalon, Lan et al. [14] employ system prompts to guide LLM agents in gameplay, while DeepRole [27] combine counterfactual regret minimization with value networks trained through self-play. For traditional Werewolf, Xu et al. [37] develop agents that generate diverse action candidates through deductive reasoning and utilize RL policies to optimize strategic gameplay. Wu et al. [36] further enhance System-2 reasoning abilities by training a Thinker module for complex logical analysis. For One Night Ultimate Werewolf (ONUW), Jin et al. [12] propose an RL-instructed language agent framework that determines appropriate discussion tactics.

Despite these advances, current SDG agents rely exclusively on textual information, ignoring non-verbal cues that humans naturally use during social interactions. In real-world settings, facial expressions and tone of voice often reveal underlying intentions, emotional states, and potential deception that complement verbal exchanges. This limitation prevents existing agents from accessing and interpreting the full spectrum of communication signals available in human gameplay.

### 2.2 Multimodal Social Interaction

Research on multimodal social interaction has expanded significantly in recent years, exploring how non-verbal cues complement verbal communication. Grauman et al. [10] present the Ego4D social benchmark for understanding social attention through video and audio. In the domain of SDGs, Lee et al. [16] introduce challenges for modeling fine-grained dynamics using densely aligned language-visual representations. Focusing specifically on ONUW, Lai et al. [13] analyze gameplay recordings to identify behavioral patterns correlating with deception.

However, these works primarily focus on retrospective analysis of social interactions rather than active participation in gameplay. They typically analyze recorded gameplay to extract patterns or identify deception markers after the fact, but do not employ these insights to inform real-time decision-making by autonomous agents. Our work differs by integrating multimodal reasoning directly into an agent's gameplay loop, enabling it to both interpret non-verbal cues from other players and strategically manage its own communication during active gameplay.



**Figure 2: The overall framework of MultiMind.** The Perceiver extracts structured information from multimodal inputs. The Reasoner and Planner iteratively optimize communication strategies: the Reasoner predicts each player’s belief states using a ToM model, while the Planner employs Monte Carlo Tree Search (MCTS) to explore possible strategies. Finally, the Actor generates natural language statements that implement the optimized strategies.

### 2.3 Theory of Mind

Theory of Mind (ToM) refers to the cognitive ability to attribute mental states to oneself and others [3], allowing agents to reason about beliefs, desires, intentions, and knowledge of other individuals. This capability is fundamental for effective social interaction and has been extensively studied in multi-agent systems. Traditional computational approaches to ToM have utilized Bayesian methods [1, 4, 15] to model beliefs and update them based on observed behaviors. More recent work has explored neural network-based approaches [2, 26, 41] that learn representations of others’ mental states directly from data. These approaches have been applied in both observational contexts, where the system infers others’ beliefs without interaction [8, 26], and interactive settings where agents must coordinate their actions [24, 32].

In the specific context of SDGs like Werewolf and ONUW, ToM reasoning is essential but has been primarily implemented at a first-order level. First-order ToM involves inferring other players’ identities, which most existing work [12, 36, 37] focuses on. However, second-order ToM reasoning about what others believe about oneself and other players remains largely unexplored in these methods. This limitation is significant because in SDGs, players must not only deduce others’ roles but also strategically manage others’ beliefs about their own identity. To address this gap, our work explicitly models second-order ToM by representing each player’s suspicion levels toward every other player, including the agent itself. This enables our agent to reason about how its communications affect others’ beliefs and to strategically select utterances that manage suspicion directed at itself.

## 3 Method

In this section, we first introduce our framework architecture and its four key components, and then present the data construction and training process of our system.

### 3.1 Overview

As shown in Figure 2, Our framework consists of four components: the Perceiver, the Reasoner, the Planner, and the Actor.

**The Perceiver** extracts essential information from dialogue history and, when playing against humans, processes facial expressions from video and emotional cues from audio.

**The Reasoner** uses the information provided by the Perceiver to infer second-order beliefs for each player, modeling how players perceive one another’s potential identities.

**The Planner** then leverages these belief states predicted by the Reasoner to determine communication strategies.

**The Actor** implements these strategies by formulating coherent statements or making decisions.

These components enable our agent to process multimodal information and engage in sophisticated ToM reasoning, allowing for more human-like social deduction capabilities.

### 3.2 Perceiver for Information Extraction

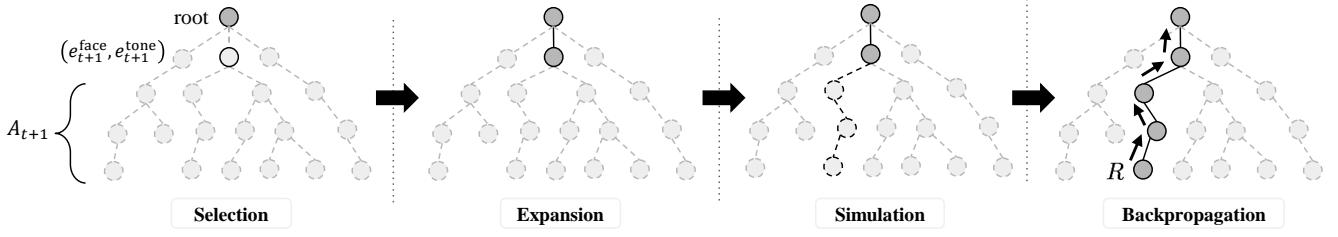
The Perceiver extracts structured information from game dialogue and, when playing against humans, processes multimodal signals including facial expressions and vocal tones.

For textual information processing, we employ an LLM to parse each player’s statements into structured triplets of  $(subject, predicate, object)$ . The subject is always the speaker, while the predicate belongs to a finite action space consisting of predefined actions: “support”, “suspect”, and “accuse as role” for each possible role. In ONUW, there are five distinct roles, resulting in a predicate space of  $2+5=7$  possible actions. The object specifies the target of the predicate, which can include the speaker themselves. Formally, we can express this extraction as:

$$A_t = f_{LLM}(S_t) \subseteq \{(p_i, a, p_j) \mid a \in \mathcal{A}, p_i, p_j \in \mathcal{P}\}, \quad (1)$$

where  $S_t$  is the statement made by player  $p_i$  at time  $t$ ,  $A_t$  represents the set of actions extracted from player  $p_i$ ’s statement,  $\mathcal{A}$  is the predicate action space, and  $\mathcal{P}$  is the set of players.

When playing against human participants, we enhance information extraction through multimodal processing. Human players are individually seated at computers, where their video and audio are recorded during gameplay. For audio processing, we utilize OSUM [7] to classify vocal tones into one of eight emotional categories (happy, sad, neutral, angry, surprise, disgust, fear, or other) and



**Figure 3: The tree structure of the MCTS algorithm in the Planner. The first level of the tree consists of nodes representing pairs of emotion labels  $(e_{t+1}^{\text{face}}, e_{t+1}^{\text{tone}})$ . The subsequent levels contain nodes that represent either action triplets  $(p_i, a, p_j)$  or “stop” actions, where multiple triplets combine to form the final action sequence  $A_{t+1}$ . Each iteration of the MCTS involves selection, expansion, and simulation, followed by backpropagation of the reward  $R$  from the terminal node back to the root.**

transcribe speech into text. Concurrently, for visual information, we employ Emotion-LLAMA [5, 6] to classify facial expressions in video frames into the same eight emotional categories. This process can be formalized as:

$$\begin{aligned} S_t, e_t^{\text{tone}} &= f_{\text{audio}}(\text{Audio}_t), \\ e_t^{\text{face}} &= f_{\text{visual}}(\text{Video}_t), \end{aligned} \quad (2)$$

where  $S_t$  is the transcribed speech, and  $e_t^{\text{tone}}$  and  $e_t^{\text{face}}$  represent the emotion labels derived from player  $p_i$ 's audio and video at time  $t$ , respectively.

The Perceiver integrates these textual actions and emotional signals to provide a comprehensive multimodal representation of each player's communication, which is then passed to the Reasoner for further processing and belief state modeling.

### 3.3 Reasoner with ToM modeling

The Reasoner processes the structured information extracted by the Perceiver and employs a specialized Theory of Mind (ToM) model to predict belief distributions representing each player's werewolf suspicions toward others.

Our ToM model adopts a Transformer [30] with causal attention where each input token represents an event in the game timeline, and each corresponding output hidden state is transformed through a linear layer to produce a belief matrix  $B$ . The causal attention mechanism ensures that the computation of each belief matrix only considers the game history up to the current time point, reflecting the progressive nature of belief formation during gameplay.

For the  $k$ -th triplet  $(p_i, a^{(k)}, p_j^{(k)}) \in A_t$  at time  $t$ , we encode it as an input token comprising five embeddings:

$$\begin{aligned} E_{t,k} = & E_{\text{subj}}(p_i) + E_{\text{pred}}(a^{(k)}) + E_{\text{obj}}(p_j^{(k)}) \\ & + E_{\text{face}}(e_t^{\text{face}}) + E_{\text{tone}}(e_t^{\text{tone}}), \end{aligned} \quad (3)$$

where  $E_{\text{subj}}$ ,  $E_{\text{pred}}$ , and  $E_{\text{obj}}$  represent embeddings for the subject, predicate, and object respectively.  $E_{\text{face}}$  encodes facial emotion labels, and  $E_{\text{tone}}$  encodes speech emotion labels. All action triplets from player  $p_i$  at time  $t$  share the same  $e_t^{\text{face}}$  and  $e_t^{\text{tone}}$ .

Let  $N_t = \sum_{\tau=1}^t |A_\tau|$  denote the total number of triplets up to time  $t$ . The complete input sequence  $E_{1:N_t}$  comprises the sequential collection of all encoded triplets:

$$E_{1:N_t} = [E_{1,1}, \dots, E_{1,|A_1|}, \dots, E_{t,1}, \dots, E_{t,|A_t|}]. \quad (4)$$

A Transformer  $\mathcal{D}$  with causal attention processes this sequence to generate hidden states:

$$h_k = \mathcal{D}(E_{1:k}), \quad \forall k \in \{1, 2, \dots, N_t\}. \quad (5)$$

The hidden state  $h_{N_t}$  is then processed through a linear layer to generate the belief matrix  $B_t$  at time  $t$ :

$$B_t[i, :] = \text{Softmax}(W_t \cdot h_{N_t} + b_t), \quad \forall i \in \{1, 2, \dots, |\mathcal{P}|\}, \quad (6)$$

where  $|\mathcal{P}|$  is the number of players. Each element  $B_t[i, j]$  denotes the probability that player  $p_i$  believes player  $p_j$  is a werewolf.

We can thus encapsulate the belief matrix computation as a function of the history of actions and emotion labels:

$$B_t = f_{\text{ToM}}(A_{1:t}, e_{1:t}^{\text{face}}, e_{1:t}^{\text{tone}}), \quad (7)$$

where  $A_{1:t}$  represents all action triplets up to time  $t$ , and  $e_{1:t}^{\text{face}}$  and  $e_{1:t}^{\text{tone}}$  represent the corresponding sequences of facial and speech emotion labels, respectively.

This ToM modeling enables our agent to reason how it is perceived by others, providing the Planner with belief state information for optimizing communication strategies.

### 3.4 Planner with Monte Carlo Tree Search

The Planner employs Monte Carlo Tree Search (MCTS) to determine optimal communication strategies based on belief states predicted by the Reasoner. Its primary objective is to generate a sequence of strategic actions  $A_{t+1}$  that minimizes the cumulative suspicion directed toward the agent by other players.

To achieve this, the Planner aims to identify a sequence of statements that will most effectively reduce the probability of being suspected as a werewolf. Let  $i$  denote the index of our agent in the player set  $\mathcal{P}$ . Formally, we define the optimization problem as:

$$(A_{t+1}^*, e_{t+1}^{\text{face}*}, e_{t+1}^{\text{tone}*}) = \arg \min_{A_{t+1}, e_{t+1}^{\text{face}}, e_{t+1}^{\text{tone}}} \sum_{j=1, j \neq i}^{|P|} B_{t+1}[j, i], \quad (8)$$

where  $B_{t+1}$  is computed using Equation (7):

$$\begin{aligned} B_{t+1} &= f_{\text{ToM}}(A_{1:t+1}, e_{1:t+1}^{\text{face}}, e_{1:t+1}^{\text{tone}}) \\ &= f_{\text{ToM}}(A_{1:t} \cup A_{t+1}, e_{1:t}^{\text{face}} \cup \{e_{t+1}^{\text{face}}\}, e_{1:t}^{\text{tone}} \cup \{e_{t+1}^{\text{tone}}\}). \end{aligned} \quad (9)$$

As shown in Figure 3, the MCTS operates on a hierarchical tree structure rooted at the current game state  $(A_{1:t}, e_{1:t}^{\text{face}}, e_{1:t}^{\text{tone}})$ . The tree's first level represents the selection of emotion labels, where each node corresponds to a specific pair  $(e_{t+1}^{\text{face}}, e_{t+1}^{\text{tone}})$ . With 8 possible  $e_{t+1}^{\text{face}}$  and 8 possible  $e_{t+1}^{\text{tone}}$ , this level contains 64 nodes.

Below the first level, the tree branches into additional levels that incrementally construct  $A_{t+1}$ . Each node at these levels represents either an action triplet  $(p_i, a, p_j)$  where  $p_i$  is always our agent, or a “stop” action that terminates the sequence. In a 5-player game, with  $|\mathcal{A}| = 7$  (as described in Section 3.2) and  $|\mathcal{P}| = 5$ , each node can branch into  $7 \times 5 + 1 = 36$  possible nodes. The maximum depth  $|A_{t+1}|_{\max}$  is constrained to 3.

For each node in the tree, we maintain two values:  $Q(n)$ , the cumulative reward at node  $n$ , and  $N(n)$ , the visit count of node  $n$ . We initialize the root node with  $Q(\text{root}) = 0$  and  $N(\text{root}) = 0$ .

The MCTS proceeds iteratively through four phases:

**Selection:** The algorithm starts at the root node and recursively traverses the tree. At each node, if the current node has unvisited children or has no children, stop and select this node. Otherwise, move to the child node  $n$  that maximizes the Upper Confidence Bound for Trees (UCT):

$$\text{UCT}(n) = \frac{Q(n)}{N(n)} + C \sqrt{\frac{\ln N(\text{parent}(n))}{N(n)}}, \quad (10)$$

where  $\text{parent}(n)$  is the parent node, and  $C$  is a constant that balances between exploitation and exploration, conventionally set to 1.414. If multiple nodes have the same maximum UCT value, one of them is randomly chosen.

**Expansion:** If the selected node has no children, the algorithm proceeds directly to the backpropagation phase. Otherwise, we randomly select one of its unvisited children, and initialize it with  $Q(n) = 0$  and  $N(n) = 0$ .

**Simulation:** From the newly initialized node, we randomly traverse down the tree. A terminal state occurs when either  $|A_{t+1}|$  reaches 3 or a “stop” action is reached.

**Backpropagation:** Once we reach a terminal state with a fully constructed action sequence  $(A_{t+1}, e_{t+1}^{\text{face}}, e_{t+1}^{\text{tone}})$ , we compute  $B_{t+1}$  using Equation (9), and calculate the reward as:

$$R = - \sum_{j=1, j \neq i}^{|P|} B_{t+1}[j, i]. \quad (11)$$

This reward represents the negative sum of suspicion towards our agent  $p_i$  from other players. We then update  $Q(n)$  and  $N(n)$  for all nodes along the path from the terminal node back to the root:

$$\begin{aligned} Q(n) &= Q(n) + R, \\ N(n) &= N(n) + 1. \end{aligned} \quad (12)$$

After 500 iterations, the algorithm selects the terminal state  $(A_{t+1}, e_{t+1}^{\text{face}}, e_{t+1}^{\text{tone}})$  with the highest  $R$ . This optimization process enables our agent to identify communication strategies that can minimize other players' suspicion toward itself.

The optimized strategies  $(A_{t+1}, e_{t+1}^{\text{face}}, e_{t+1}^{\text{tone}})$  determined by the Planner are then passed to the Actor, which transforms these strategic actions into natural language utterances.

### 3.5 Actor for Response Generation

The Actor transforms the communication strategies determined by the Planner into coherent natural language utterances that can be effectively communicated during gameplay.

For the action sequence  $A_{t+1}$ , the Actor employs an LLM to generate a cohesive statement that expresses these actions in natural language. This process can be formalized as:

$$S_{t+1} = g_{\text{LLM}}(\mathcal{R}, G_t, S_{1:t}, e_{1:t}^{\text{face}}, e_{1:t}^{\text{tone}}, A_{t+1}), \quad (13)$$

where  $S_{t+1}$  is the agent's statement,  $\mathcal{R}$  represents the game rules,  $G_t$  is the game state at time  $t$ ,  $S_{1:t}$  denotes the dialogue history, and  $e_{1:t}^{\text{face}}$  and  $e_{1:t}^{\text{tone}}$  are the historical facial and speech emotion labels.

For the emotion labels  $e_{t+1}^{\text{face}}$  and  $e_{t+1}^{\text{tone}}$ , the Actor presents them directly as text labels alongside the generated statement. These emotion indicators help human players and other agents interpret the agent's emotional state, adding another dimension to the communication beyond the textual content.

### 3.6 Training Process

In our framework, the only component requiring training is the ToM model in the Reasoner. We employ a two-phase training approach: first utilizing data from agent self-play, followed by fine-tuning with human gameplay data from Lai et al. [13].

**3.6.1 Self-Play Training.** For the initial training, we generate a dataset through agent self-play in 5-player ONUW games. Each agent is randomly assigned one of three LLM backends: GPT-4o [22], Qwen2.5-14B-Instruct [25], or Llama-3.1-8B-Instruct [9]. Since the ToM model is not yet available, we modify Equation (13) to:

$$S_{t+1}, e_{t+1}^{\text{face}}, e_{t+1}^{\text{tone}} = g_{\text{LLM}}(\mathcal{R}, G_t, S_{1:t}, e_{1:t}^{\text{face}}, e_{1:t}^{\text{tone}}). \quad (14)$$

This means that each LLM generates statements and simultaneously selects emotional labels from the predefined sets of 8 facial expressions and 8 vocal tones.

After each agent's speech, we prompt all agents to identify their werewolf suspicions. Let  $S_i$  be the set of players suspected by agent  $p_i$ , and  $|S_i|$  be the number of suspected players. We construct the ground truth belief matrix  $B_t^{\text{GT}}[i, j]$  as:

$$B_t^{\text{GT}}[i, j] = \begin{cases} \frac{1}{|S_i|} & \text{if } |S_i| > 0 \text{ and } p_j \in S_i, \\ \frac{1}{|\mathcal{P}|} & \text{if } |S_i| = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

where  $|\mathcal{P}|$  is the number of all players in the game.

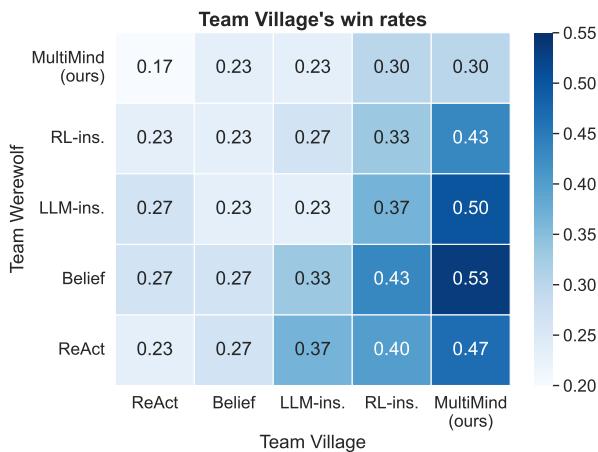
We train the ToM model using a cross-entropy loss between the predicted belief matrices and the ground truth matrices:

$$\mathcal{L} = - \sum_{t=1}^T \sum_{i=1}^{|\mathcal{P}|} \sum_{j=1}^{|\mathcal{P}|} B_t^{\text{GT}}[i, j] \log(B_t^{\text{pred}}[i, j]). \quad (16)$$

**3.6.2 Human Data Fine-Tuning.** Following self-play training, we fine-tune our ToM model using the human gameplay dataset from Lai et al. [13]. This dataset comprises 163 videos containing timestamped speech transcriptions and player intention annotations. Each video contains between 1-3 games of 5-player ONUW, totaling 199 complete games.

**Table 1: Performance of agents in mixed-agent games.** “Participations” counts how many times an agent is selected across 400 games (one agent can be selected multiple times in a single game). “Avg. Votes” denotes the average number of votes received by the agent per game (maximum 5 votes).

Agent	In Team Werewolf			In Team Village			Overall
	Participations	Avg. Votes (↓)	Win Rate (↑)	Participations	Avg. Votes (↓)	Win Rate (↑)	Win Rate (↑)
ReAct	79	1.29	58.2	339	1.06	33.0	37.8
Belief	66	1.44	59.1	315	0.97	35.6	39.6
LLM-instructed	82	1.20	67.1	318	0.98	35.9	42.3
RL-instructed	87	1.32	57.5	290	<b>0.83</b>	37.2	41.9
MultiMind (ours)	86	<b>1.05</b>	<b>70.9</b>	338	<b>0.83</b>	<b>44.4</b>	<b>49.8</b>



**Figure 4: Team Village’s win rates.** We conduct 30 games for each setting and calculate the win rates.

For each video, we extract individual speech segments using the provided timestamps. To obtain  $e_t^{\text{tone}}$ , we extract the audio from each speech segment and classify it using OSUM. To obtain  $e_t^{\text{face}}$ , we leverage Emotion-LLaMA. However, Emotion-LLaMA can only process single-person videos, while the original recordings capture multiple players simultaneously. Fortunately, each player’s position remains relatively stable throughout the gameplay as participants are seated. Therefore, we manually crop regions corresponding to each player. These individual player video segments are then processed by Emotion-LLaMA to obtain  $e_t^{\text{face}}$ .

To construct  $B_t^{\text{GT}}$ , we utilize the player intention annotations available in the dataset. For segments lacking annotations, we employ GPT-4o by providing it with the complete game record and existing intention annotations, prompting it to infer players’ suspicions at specific timestamps.

## 4 Experiments

In this section, we first present the results of our agent competing against other baselines, then conduct an ablation study of the ToM model and MCTS, showcase a user study evaluating our agent’s

performance against human players, and finally discuss the contribution of multimodal cues in ToM modeling.

### 4.1 Implementation Details

**4.1.1 Training Scheme.** Deploying Qwen2.5-14B-Instruct and Llama-3.1-8B-Instruct on a single A800 GPU to complete one game requires an average of 3.5 minutes. We utilized 4 A800 GPUs to generate 5,300 games, which took approximately 77 hours. From this dataset, 5,000 games were allocated to the training set and 300 games to the validation set. We implemented early stopping based on the validation loss, terminating training when the validation loss began to increase. The ToM model, comprising 25.4M parameters (with a hidden size of 512 and 8 decoder-only layers), was trained on a single A800 GPU for 80 epochs with a batch size of 32 and a learning rate of 5e-5, completing in 4.5 hours.

From the 199 games in the human dataset, we allocated 184 games for our training set and 15 games for our validation set. The fine-tuning process for the ToM model was conducted on a single A800 GPU. We employed a batch size of 32 and a learning rate of 5e-5, running for 35 epochs. The entire fine-tuning procedure was completed in 11 minutes.

**4.1.2 Evaluation Setup.** We conduct our experiments using the 5-player ONUW environment<sup>1</sup> implemented by Jin et al. [12]. This implementation offers two difficulty levels (easy and hard), and we use the hard setting for our experiments. The game includes 1 werewolf, 1 seer, 1 robber, 1 troublemaker, and 1 insomniac. The werewolf belongs to Team Werewolf, while all other roles belong to Team Village. During the daytime phase, all players engage in three rounds of discussion, followed by a voting process. If the werewolf is among those with the highest votes, Team Village wins. Otherwise, Team Werewolf wins.

For comparison, we implement the ReAct [39] agent, which directly prompts the LLM with raw observations to generate its reasoning and actions. Additionally, we include three agent variants in Jin et al. [12]: Belief, LLM-instructed, and RL-instructed.

For agent players,  $e_t^{\text{face}}$  and  $e_t^{\text{tone}}$  appear only as text labels generated by the agents themselves. For human players, each human participant is seated individually at a computer where their video and audio are recorded. We then use the methods described in Section 3.2 to extract  $S_t$ ,  $e_{1:t}^{\text{face}}$ , and  $e_{1:t}^{\text{tone}}$ .

<sup>1</sup><https://github.com/KylJin/Werewolf>

**Table 2: Comparison of Reasoner variants.** Each variant plays against ReAct agent as either Team Werewolf or Team Village. 50 games are conducted for each setting. “Avg. Votes” indicates the average number of votes received by each agent per game (maximum 5 votes). For the ToM model, we use a single 3090 GPU with batch size 1. For Gemini-2.0-Flash, we use 5 threads to simultaneously predict each row of  $B_t$ .

Reasoner	MCTS iter	Planning Time (s)	Team Werewolf		Team Village		Overall
			Avg. Votes (↓)	Win Rate (↑)	Avg. Votes (↓)	Win Rate (↑)	
ToM model	1000	18.6	<b>0.91</b>	<b>84.0</b>	<b>0.75</b>	<b>50.0</b>	<b>67.0</b>
ToM model	500	7.8	0.92	<b>84.0</b>	0.77	48.0	66.0
ToM model	200	4.4	1.12	76.0	0.88	26.0	51.0
Gemini-2.0-Flash	200	634.2	1.09	78.0	0.89	28.0	53.0

**Table 3: Comparison of Planner variants.** Each variant plays against the ReAct agent as either Team Werewolf or Team Village. For MCTS, we set 500 iterations, which means the ToM model is called 500 times to calculate rewards. For other variants, we traverse 500 nodes, also calling the ToM model 500 times to calculate rewards, and select the path to the highest reward node as the final strategies. We conduct 50 games for each setting.

Planner	Team Werewolf		Team Village	Overall
	Win Rate (↑)	Win Rate (↑)	Win Rate (↑)	
Random	76.0	38.0	57.0	
DFS	68.0	20.0	44.0	
BFS	80.0	42.0	61.0	
MCTS (Ours)	<b>84.0</b>	<b>48.0</b>	<b>66.0</b>	

During the training of our ToM model, we utilized data generated by GPT-4o, Qwen2.5-14B-Instruct, and Llama-3.1-8B-Instruct. Therefore, to demonstrate our generalizability, all experiments use Gemini-2.0-Flash as the backend LLM unless otherwise specified.

When playing against other agents, we use the ToM model trained on self-play data as described in Section 3.6, while for games involving human players, we employ the human-data fine-tuned version of the ToM model.

## 4.2 Main Results

Following Jin et al. [12], we evaluate our agent’s performance when playing as either Team Village or Team Werewolf against baseline agents. The win rates for Team Village are presented in Figure 4. The rows of the matrix represent the agent type used by Team Village and the columns represent the agent type used by Team Werewolf. In each game, all players on a given team (Village or Werewolf) use the same agent type.

As observed across each row, our agent consistently achieves the highest win rates when deployed as Team Village. Similarly, each column reveals that our agent poses significant challenge when playing as Team Werewolf, as evidenced by the low win rates for Team Village when facing our agent.

Notably, our agent’s performance boost is more significant when playing as Team Village than as Team Werewolf. We attribute this

**Table 4: Performance against human players.** “Avg. Votes” indicates the average number of votes each agent received per game (maximum 5 votes), and “Avg. Human Votes” represents the average number of votes each agent received from human players per game (maximum 1 vote).

Agent	Avg. Votes (↓)	Avg. Human Votes (↓)	Win Rate (↑)
Human	0.90	—	40.0
ReAct	1.18	0.32	26.2
Belief	1.16	0.28	34.8
LLM-instructed	1.03	0.24	33.8
RL-instructed	<b>0.87</b>	0.22	40.6
MultiMind	<b>0.87</b>	<b>0.19</b>	<b>42.2</b>

discrepancy to the experimental setup where all players on a team share identical agent types. This homogeneous team structure particularly amplifies Team Village’s advantage by enabling strong coordination through consistent reasoning patterns.

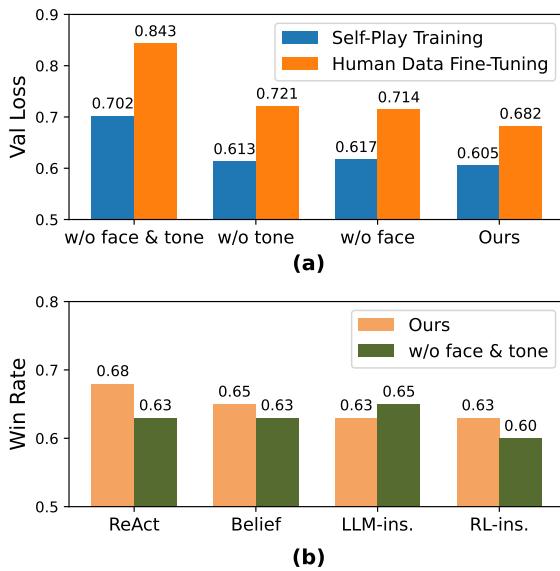
To further evaluate our agent’s performance in more realistic scenarios, we conduct 400 games with random agent selection. For each game, the 5 players are randomly selected from our agent and four baseline agents. This setup allows us to assess how each agent performs when integrated into heterogeneous teams, mimicking real-world gameplay with diverse player strategies.

Table 1 presents the results. MultiMind achieves the highest win rate in both Team Werewolf (70.9%) and Team Village (44.4%), with the lowest average votes against it across all scenarios. This demonstrates our agent’s superior suspicion avoidance, particularly when acting as a werewolf.

## 4.3 Ablation Study

To evaluate the contribution of individual components within our framework, we conduct an ablation study focused on the Reasoner and the Planner, which are highly coupled in our architecture.

For the Reasoner, our primary approach uses a lightweight ToM model to predict belief states, allowing the Planner to explore numerous strategies efficiently. As an alternative, we implement a version that directly employs the backend LLM to infer other players’ beliefs, modifying Equation (7) to:



**Figure 5: Ablation study on multimodal inputs. (a) Comparison of ToM model validation loss. (b) Comparison of win rates. Our agent plays 40 games against each baseline, with 20 games as Team Werewolf and 20 games as Team Village.**

$$B_t = f_{LLM}(\mathcal{R}, G_t, S_{1:t}, e_{1:t}^{\text{face}}, e_{1:t}^{\text{tone}}), \quad (17)$$

where  $\mathcal{R}$  represents the game rules,  $G_t$  is the game state at time  $t$ , and  $S_{1:t}$  is the complete dialogue history up to time  $t$ .

Since the LLM-based Reasoner requires significantly more computation time than our ToM model, we change the number of MCTS iterations in the Planner, ensuring a fair comparison. We evaluate all variants against the ReAct agent, with our agent playing as either Team Village or Team Werewolf. The results in Table 2 demonstrate that with the same number of MCTS iterations, our ToM model achieves comparable performance to the LLM-based Reasoner with significantly lower computational cost. This enables us to increase the number of MCTS iterations, spending more time on planning to further improve performance.

The experiments also reveal diminishing returns as MCTS iteration count increases. The performance gap between 200 and 500 iterations is substantial, with overall win rate improving from 51.0% to 66.0%. However, increasing from 500 to 1000 iterations yields only negligible improvement.

To demonstrate the contribution of our MCTS Planner, we compare it against alternative planning strategies: random sampling, depth-first search (DFS), and breadth-first search (BFS). Table 3 presents the results of this comparison. All variants use our ToM model as the Reasoner with identical computational budgets. The MCTS Planner outperforms all alternatives across both Team Village and Team Werewolf scenarios. Notably, DFS performs worse than random sampling due to its tendency to get trapped in suboptimal exploration paths, while BFS achieves relatively good results by providing a more balanced exploration strategy.

#### 4.4 User Study

To evaluate our agent's performance in a more realistic setting with human players, we recruited 8 volunteers to participate in our study. After explaining the game rules to these participants, each human player completed 10 games of 5-player ONUW. Each game consisted of 1 human player alongside 4 AI players randomly selected from 5 agents (MultiMind, ReAct, Belief, LLM-instructed, and RL-instructed).

We measured both the win rate of each agent and the number of votes they received. Table 4 presents these results. MultiMind achieves the highest win rate while receiving the lowest number of human votes. This aligns with our design objective of suspicion minimization through ToM reasoning.

Notably, while MultiMind still performs well in terms of overall votes received, its advantage is less pronounced compared to our agent-only experiments. We attribute this to the human-data fine-tuned ToM model used in these games. While specializing in human behavior patterns improves deception against humans, it may slightly reduce effectiveness against agent opponents compared to the self-play trained version.

#### 4.5 Multimodal Cues in ToM Modeling

To evaluate the impact of multimodal information in ToM modeling, specifically how  $e_t^{\text{face}}$  and  $e_t^{\text{tone}}$  influence the prediction of  $B_t$ , we conduct an ablation experiment. We separately remove the  $E_{\text{face}}(e_t^{\text{face}})$  and  $E_{\text{tone}}(e_t^{\text{tone}})$  terms from Equation (3), then train the ToM model using the same data and parameters described in Section 3.6. To compare performance, we measure the lowest validation loss achieved during both self-play training and human data fine-tuning phases. In both phases, training is stopped when validation loss begins to increase.

As shown in Figure 5(a), both  $e_t^{\text{face}}$  and  $e_t^{\text{tone}}$  in Equation (3) contribute to the model's ability to predict belief states. The complete model incorporating both modalities achieve lower validation loss compared to ablated variants. This demonstrates that multimodal cues provide valuable additional context for modeling players' mental states in social deduction games. Besides, Figure 5(b) depicts the win rates of our agent against other baselines when multimodal information is not used. While there is a slight decrease in performance, the win rates still exceed 0.6 against all baselines.

## 5 Conclusion

In this paper, we introduced MultiMind, a novel framework that enhances LLM agents for social deduction games by integrating multimodal reasoning and ToM capabilities. Through a combination of ToM reasoning and MCTS planning, MultiMind effectively optimizes communication strategies to minimize suspicion from other players, demonstrating significant improvements in gameplay performance. Our work highlights the potential of multimodal cues in complex social reasoning, paving the way for the development of AI agents capable of sophisticated social interactions.

## Acknowledgments

This research is supported by Tencent Rhino-Bird Focused Research Program, National Natural Science Foundation of China

(No. 62406267), and the Guangzhou Municipal Science and Technology Project (No. 2025A04J4070).

## References

- [1] Chris L. Baker, Rebecca Saxe, and Joshua B. Tenenbaum. 2009. Action understanding as inverse planning. *Cognition* 113, 3 (2009), 329–349. doi:10.1016/j.cognition.2009.07.005 Reinforcement learning and higher cognition.
- [2] Cristian-Paul Bara, Sky CH-Wang, and Joyce Chai. 2021. MindCraft: Theory of Mind Modeling for Situated Dialogue in Collaborative Tasks. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 1112–1125. doi:10.18653/v1/2021.emnlp-main.85
- [3] Michael Bratman. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press, Cambridge.
- [4] Moritz C. Buehler and Thomas H. Weisswange. 2020. Theory of Mind based Communication for Human Agent Cooperation. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, 1–6. doi:10.1109/ICHMS49158.2020.9209472
- [5] Zebang Cheng, Zhi-Qi Cheng, Jun-Yan He, Kai Wang, Yuxiang Lin, Zheng Lian, Xiaojiang Peng, and Alexander Hauptmann. 2024. Emotion-LLaMA: Multimodal Emotion Recognition and Reasoning with Instruction Tuning. In *Advances in Neural Information Processing Systems*, A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomeczak, and C. Zhang (Eds.), Vol. 37. Curran Associates, Inc., 110805–110853. [https://proceedings.neurips.cc/paper\\_files/paper/2024/file/c7f43ada17acc234f568dc66da527418-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2024/file/c7f43ada17acc234f568dc66da527418-Paper-Conference.pdf)
- [6] Zebang Cheng, Shuyuan Tu, Dawei Huang, Minghan Li, Xiaojiang Peng, Zhi-Qi Cheng, and Alexander G. Hauptmann. 2024. SZTU-CMU at MER2024: Improving Emotion-LLaMA with Conv-Attention for Multimodal Emotion Recognition. In *Proceedings of the 2nd International Workshop on Multimodal and Responsible Affective Computing* (Melbourne VIC, Australia) (MRAC '24). Association for Computing Machinery, New York, NY, USA, 78–87. doi:10.1145/3689092.3689404
- [7] Xuelong Geng, Kun Wei, Qijie Shao, Shuiyun Liu, Zhennan Lin, Zhixian Zhao, Guojian Li, Wenjie Tian, Peikun Chen, Yangze Li, Pengcheng Guo, Mingchen Shao, Shuiyun Wang, Yuang Cao, Chengyu Wang, Tianyi Xu, Yuhang Dai, Xinfu Zhu, Yue Li, Li Zhang, and Lei Xie. 2025. OSUM: Advancing Open Speech Understanding Models with Limited Resources in Academia. *arXiv preprint arXiv:2501.13306* (2025).
- [8] Erin Grant, Aida Nematzadeh, and Thomas L. Griffiths. 2017. How Can Memory-Augmented Neural Networks Pass a False-Belief Task? *Cognitive Science* (2017). <https://api.semanticscholar.org/CorpusID:7340345>
- [9] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archis Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Greggerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, Danny Wyatt, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, et al. 2024. The Llama 3 Herd of Models. *arXiv:2407.21783 [cs.AI]* <https://arxiv.org/abs/2407.21783>
- [10] Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, Miguel Martin, Tushar Nagarajan, Ilijा Radosavović, Santhosh Kumar Ramakrishnan, Fiona Ryan, Jayant Sharma, Michael Wray, Mengmeng Xu, Eric Zhongcong Xu, Chen Zhao, Siddhant Bansal, Dhruv Batra, Vincent Cartillier, Sean Crane, Tien Do, Morrie Doulaty, Akshay Erapalli, Christoph Feichtenhofer, Adriano Fragnemeni, Qichen Fu, Abrham Gebreselasie, Cristina González, James Hillis, Xuhua Huang, Yifei Huang, Wenqi Jia, Weslie Khoo, et al. 2022. Ego4D: Around the World in 3,000 Hours of Egocentric Video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 18995–19012.
- [11] Yuya Hirata, Michimasa Inaba, Kenichi Takahashi, Fujio Toriumi, Hirotaka Osawa, Daisuke Katagami, and Kousuke Shinoda. 2016. Werewolf Game Modeling Using Action Probabilities Based on Play Log Analysis. In *Computers and Games*. <https://api.semanticscholar.org/CorpusID:37838481>
- [12] Xuanfa Jin, Ziyian Wang, Yali Du, Meng Fang, Haifeng Zhang, and Jun Wang. 2024. Learning to Discuss Strategically: A Case Study on One Night Ultimate Werewolf. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. <https://openreview.net/forum?id=1f82rnwCbl>
- [13] Bolin Lai, Hongxin Zhang, Miao Liu, Aryan Pariani, Fiona Ryan, Wenqi Jia, Shirley Anugrah Hayati, James Rehg, and Diyi Yang. 2023. Werewolf Among Us: Multimodal Resources for Modeling Persuasion Behaviors in Social Deduction Games. In *Findings of the Association for Computational Linguistics: ACL 2023*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 6570–6588. doi:10.18653/v1/2023.findings-acl.411
- [14] Yihuai Lan, Zhiqiang Hu, Lei Wang, Yang Wang, Deheng Ye, Peilin Zhao, Ee-Peng Lim, Hui Xiong, and Hao Wang. 2024. LLM-Based Agent Society Investigation: Collaboration and Confrontation in Avalon Gameplay. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen (Eds.). Association for Computational Linguistics, Miami, Florida, USA, 128–145. doi:10.18653/v1/2024.emnlp-main.7
- [15] Jin Joo Lee, Fei Sha, and Cynthia Breazeal. 2019. A Bayesian Theory of Mind Approach to Nonverbal Communication. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 487–496. doi:10.1109/HRI.2019.8673023
- [16] Sangmin Lee, Bolin Lai, Fiona Ryan, Bikram Boote, and James M. Rehg. 2024. Modeling Multimodal Social Interactions: New Challenges and Baselines with Densely Aligned Representations. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14585–14595. doi:10.1109/CVPR52733.2024.01382
- [17] Zaijing Li, Yuqian Xie, Rui Shao, Gongwei Chen, Dongmei Jiang, and Liqiang Nie. 2024. Optimus-1: Hybrid Multimodal Memory Empowered Agents Excel in Long-Horizon Tasks. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. <https://openreview.net/forum?id=XXOMCwZ6by>
- [18] Shano Liang, Max Chen, Phoebe O. Troups Dugas, Gillian Smith, and Rose Bohrer. 2025. The Collaborative Sensemaking Play of Jubensha Games: A Deconstruction, Taxonomy, and Analysis. *ACM Games* 3, 1, Article 6 (March 2025), 34 pages. doi:10.1145/3721121
- [19] Jonathan Light, Min Cai, Sheng Shen, and Ziniu Hu. 2023. From Text to Tactic: Evaluating LLMs Playing the Game of Avalon. In *NeurIPS 2023 Foundation Models for Decision Making Workshop*. <https://openreview.net/forum?id=ltUrSryS0K>
- [20] Weiyu Ma, Qirui Mi, Yongcheng Zeng, Xue Yan, Runji Lin, Yuqiao Wu, Jun Wang, and Haifeng Zhang. 2024. Large Language Models Play StarCraft II: Benchmarks and A Chain of Summarization Approach. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. <https://openreview.net/forum?id=kEPpD7yETM>
- [21] Noritsugu Nakamura, Michimasa Inaba, Kenichi Takahashi, Fujio Toriumi, Hirotaka Osawa, Daisuke Katagami, and Kousuke Shinoda. 2016. Constructing a Human-like agent for the Werewolf Game using a psychological model based multiple perspectives. *2016 IEEE Symposium Series on Computational Intelligence (SSCI)* (2016), 1–8. <https://api.semanticscholar.org/CorpusID:34482956>
- [22] OpenAI, ;, Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Mądry, Alex Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, Alex Nichol, Alex Paino, Alex Renzin, Alex Tachard Passos, Alexander Kirillov, Alexi Christakis, Alexia Conneau, Ali Kamali, Allan Jabri, Allison Moyer, Allison Tam, Amadou Crookes, Amin Tootoochian, Amin Tootoochian, Ananya Kumar, Andrei Vallone, Andrej Karpathy, Andrew Braunestein, Andrew Cann, Andrew Codispoti, Andrew Galu, Andrew Kondrich, Andrew Tulloch, Andrey Mishchenko, Angela Baek, Angela Jiang, Antoine Pelisse, Antonio Woodford, Anuj Gosalia, Arka Dhar, et al. 2024. GPT-4o System Card. *arXiv:2410.21276 [cs.CL]* <https://arxiv.org/abs/2410.21276>
- [23] Yiran Qin, Enshen Zhou, Qichang Liu, Zhenfei Yin, Lu Sheng, Ruimao Zhang, Yu Qiao, and Jing Shao. 2024. MP5: A Multi-modal Open-ended Embodied System in Minecraft via Active Perception. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16307–16316. doi:10.1109/CVPR52733.2024.01543
- [24] Liang Qiu, Yizhou Zhao, Yuan Liang, Pan Lu, Weiyan Shi, Zhou Yu, and Song-Chun Zhu. 2022. Towards Socially Intelligent Agents with Mental State Transition and Human Value. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Oliver Lemon, Dilek Hakkani-Tur, Junyi Jessy Li, Arash Ashrafzadeh, Daniel Hernández García, Malike Alikhani, David Vandyke, and Ondřej Dušek (Eds.). Association for Computational Linguistics, Edinburgh, UK, 146–158. doi:10.18653/v1/2022.sigdial-1.16
- [25] Qwen, ;, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Kemeng Lu, Kepin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihang Qiu. 2025. Qwen.25 Technical Report. *arXiv:2412.15115 [cs.CL]* <https://arxiv.org/abs/2412.15115>
- [26] Neil Rabinowitz, Frank Perbet, Francis Song, Chiyan Zhang, S. M. Ali Eslami, and Matthew Botvinick. 2018. Machine Theory of Mind. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 80)*, Jennifer Dy and Andreas Krause (Eds.). PMLR, 4218–4227. <https://proceedings.mlr.press/v80/rabinowitz18a.html>
- [27] Jack Serrino, Max Kleiman-Weiner, David C. Parkes, and Joshua B. Tenenbaum. 2019. *Finding friend and foe in multi-agent games*. Curran Associates Inc., Red Hook, NY, USA.
- [28] Xiao Shao, Weifu Jiang, Fei Zuo, and Mengqing Liu. 2024. SwarmBrain: Embodied agent for real-time strategy game StarCraft II via large language models. *arXiv:2401.17749 [cs.AI]* <https://arxiv.org/abs/2401.17749>

- [29] Zijing Shi, Meng Fang, Shunfeng Zheng, Shilong Deng, Ling Chen, and Yali Du. 2023. Cooperation on the Fly: Exploring Language Agents for Ad Hoc Teamwork in the Avalon Game. arXiv:2312.17515 [cs.CL] <https://arxiv.org/abs/2312.17515>
- [30] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. [https://proceedings.neurips.cc/paper\\_files/paper\\_2017/file/3f5ee243547dee91fdb053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper_2017/file/3f5ee243547dee91fdb053c1c4a845aa-Paper.pdf)
- [31] Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. 2024. Voyager: An Open-Ended Embodied Agent with Large Language Models. *Transactions on Machine Learning Research* (2024). <https://openreview.net/forum?id=ehfRiF0R3a>
- [32] Qiaosi Wang, Koustuv Saha, Eric Gregori, David Joyner, and Ashok Goel. 2021. Towards Mutual Theory of Mind in Human-AI Interaction: How Language Reflects What Students Perceive About a Virtual Teaching Assistant. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 384, 14 pages. doi:10.1145/3411764.3445645
- [33] Shenzhi Wang, Chang Liu, Zilong Zheng, Siyuan Qi, Shuo Chen, Qisen Yang, Andrew Zhao, Chaofei Wang, Shiji Song, and Gao Huang. 2023. Avalon's Game of Thoughts: Battle Against Deception through Recursive Contemplation. arXiv:2310.01320 [cs.AI] <https://arxiv.org/abs/2310.01320>
- [34] Tianhe Wang and Tomoyuki Kaneko. 2018. Application of Deep Reinforcement Learning in Werewolf Game Agents. *2018 Conference on Technologies and Applications of Artificial Intelligence (TAAI)* (2018), 28–33. <https://api.semanticscholar.org/CorpusID:57191228>
- [35] Dekun Wu, Haochen Shi, Zhiyuan Sun, and Bang Liu. 2024. Deciphering Digital Detectives: Understanding LLM Behaviors and Capabilities in Multi-Agent Mystery Games. In *Findings of the Association for Computational Linguistics: ACL 2024*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand, 8225–8291. doi:10.18653/v1/2024.findings-acl.490
- [36] Shuang Wu, Liwen Zhu, Tao Yang, Shiwei Xu, Qiang Fu, Yang Wei, and Haobo Fu. 2024. Enhance Reasoning for Large Language Models in the Game Werewolf. arXiv:2402.02330 [cs.AI] <https://arxiv.org/abs/2402.02330>
- [37] Yuzhuang Xu, Shuo Wang, Peng Li, Fuwen Luo, Xiaolong Wang, Weidong Liu, and Yang Liu. 2023. Exploring large language models for communication games: An empirical study on werewolf. *arXiv preprint arXiv:2309.04658* (2023).
- [38] Zelai Xu, Chao Yu, Fei Fang, Yu Wang, and Yi Wu. 2024. Language agents with reinforcement learning for strategic play in the Werewolf game. In *Proceedings of the 41st International Conference on Machine Learning* (Vienna, Austria) (ICML '24). JMLR.org, Article 2285, 31 pages.
- [39] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. ReAct: Synergizing Reasoning and Acting in Language Models. In *International Conference on Learning Representations (ICLR)*.
- [40] Zheng Zhang, Yihua Lan, Yangsen Chen, Lei Wang, Xiang Wang, and Hao Wang. 2025. DVM: Towards Controllable LLM Agents in Social Deduction Games. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1–5. doi:10.1109/ICASSP49660.2025.10888525
- [41] Pei Zhou, Andrew Zhu, Jennifer Hu, Jay Pujara, Xiang Ren, Chris Callison-Burch, Yejin Choi, and Prithviraj Ammanabrolu. 2023. I Cast Detect Thoughts: Learning to Converse and Guide with Intents and Theory-of-Mind in Dungeons and Dragons. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Anna Rogers, Jordan Boyd-Graber, and Naoki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 11136–11155. doi:10.18653/v1/2023.acl-long.624