# Enhance Reasoning for Large Language Models in the Game Werewolf

2024.02.14 Lab Seminar

신민정

# Enhance Reasoning for Large Language Models in the Game Werewolf

**Shuang Wu** [1 *] **Liwen Zhu** [1 *] **Tao Yang** [1] **Shiwei Xu** [1] **Qiang Fu** [1] **Wei Yang** [1] **Haobo Fu** [1]

[*]Equal contribution [1]Tencent AI Lab, Shenzhen, China. Correspondence to: Shuang Wu <shawnswu@tencent.com>.

Preliminary work. Under review.

# Abstract


문제

- LLM 기반 에이전트는 복잡한 논리 분석, 도메인 특화 지식이 필요한 고급 추론 작업을 수행하는 데 한계가 있음

방법

- Dual-process theory를 기반으로 추론 계층 구조 형성(System-1, System-2)
- Listener(LLM) + Thinker(외부 module) + Presenter(LLM) 구조
- LLM과 Thinker 사이의 커뮤니케이션 프로토콜 개발
- 9-player Werewolf 게임으로 테스트

결과

- 추론, 언어 생성, 온라인 게임 평가에서 새로운 framework이 우수한 성능을 보임
- 6B LLM은 GPT4보다 나은 결과 달성

의의

- LLM 기반 에이전트의 Reasoning 능력 향상
- Social Deduction Game 데이터셋 제작

# Background

## LLM의 한계

- demonstration에 그침
- practical solution을 내놓지 못함
- higher-level reasoning, planning 부족
- model의 generality를 유지해야 해 fine-tuning에도 한계가 있음

## Related Work

1. Enhancing Reasoning in LLMs
   - LLM+P (Liu et al., 2023a)
   - RAG (Lewis et al., 2020)
   - Galactica (Taylor et al., 2022) …
2. AI for Social Deduction Games
   - DeepRole (Serrino et al., 2019)
   - Deep Wolf (Shibata et al., 2023) …

# Methods - Data Preparation (Werewolf Game)

Werewolves Witch Hunter Seer Villagers


Dataset

- 관전자 모드로 18,800 게임 세션 녹화
  - 플레이 7,000 시간, speech 6,000시간
- Web-crawling, OCR-processed Werewolf literature
- Paraformer model 활용, Automatic Speech recognition fine-tuning

# Methods

System-2

System-1

- Structured features
- Prompt instructions

## *Thinker*

- Cognitive core
- 계획 수립
- deep logical analysis, domain-specific knowledge

- Strategic instructions

## *Listener*

- 자연어 처리
- 입력 데이터를 Thinker 가 해석할 수 있는 포맷으로 변경

## *Presenter*

- 현재 상황 + 전략 → 출력 텍스트 생성
- 텍스트의 논리성, 일관성, 사실성 확인

# Methods - Framework

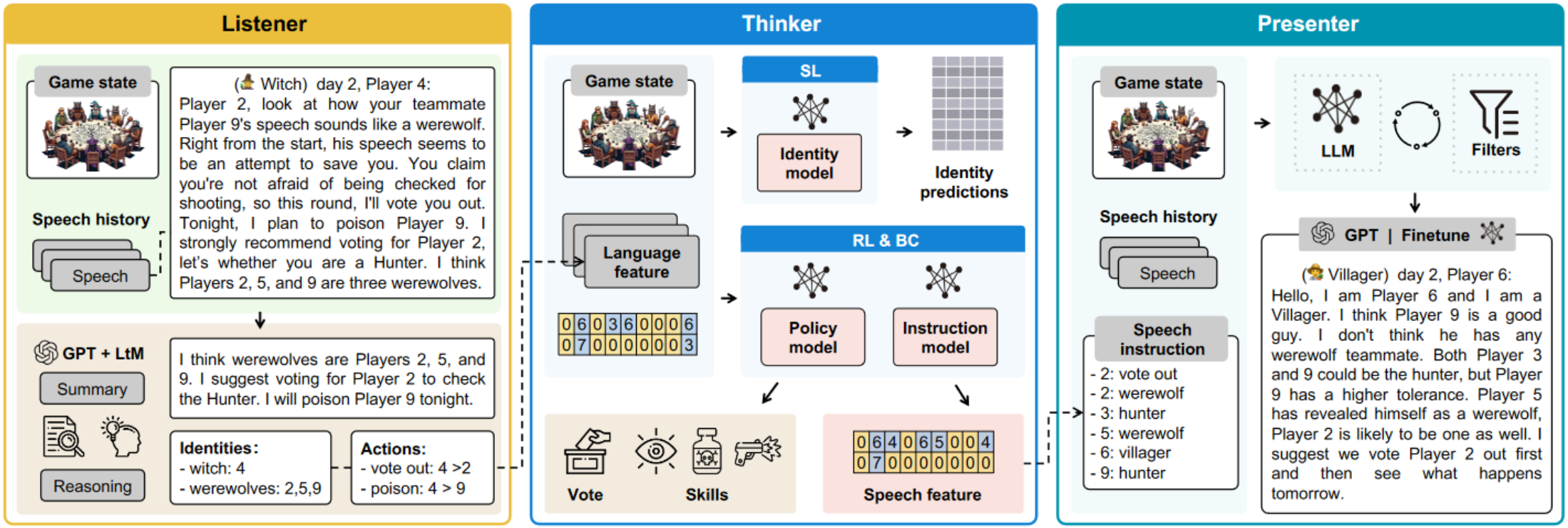


Figure 1: Overall processing framework and modules in the Werewolf implementation.

# Methods

System-2

## *Thinker*

- Direct knowledge from databases + various optimizing techniques
- Imitation Learning (Behavior Cloning) + Reinforcement Learning
- GAN + population based training + fictitious self-play

$$\mathcal{L}_{\text{BC}}(\theta) = -\mathbb{E}_{s,a\sim\mathcal{D}}[\log \pi_\theta(a|s)]$$

$$\mathcal{L}_{\text{RL}}(\theta) = -\mathbb{E}_{s,a\sim\pi_{\theta'}}\left[\frac{\pi_\theta(a|s)}{\pi_{\theta'}(a|s)} A^{\pi_\theta}(s,a)\right]$$

$$\mathcal{L} = \alpha\mathcal{L}_{\text{BC}}(\theta) + \mathcal{L}_{\text{RL}}(\theta) + \beta\mathcal{L}_{\text{id}}(\phi)$$

System-1

## *Listener*

- Synthesize & summarize (LtM prompting)
- Reasoning and feature extraction
- + ChatGLM-6B fine-tune

$$\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_N]^T$$

- Extract language features
- Compare for similarity to original speech instructions

## *Presenter*

- Controllability (strategic instructions from the *Thinker*)
- Quality (no hallucination)
- + ChatGLM-6B fine-tune

# Experiments - (1) Deductive Reasoning

목표

- Reasoning 능력 측정

방법

- Dataset에서 300 게임 선정
- Villager의 입장에서 특수 역할과 Werewolf identify

실험 대상

GPT 3.5 | GPT 3.5 + LtM | Thinker

GPT 4 | GPT 4 + LtM | Human

gpt-35-turbo-16k 0613
gpt-4 1106-Preview

* 데이터셋의 인간은 werewolf에만 투표

# Experiments - (1) Deductive Reasoning Results

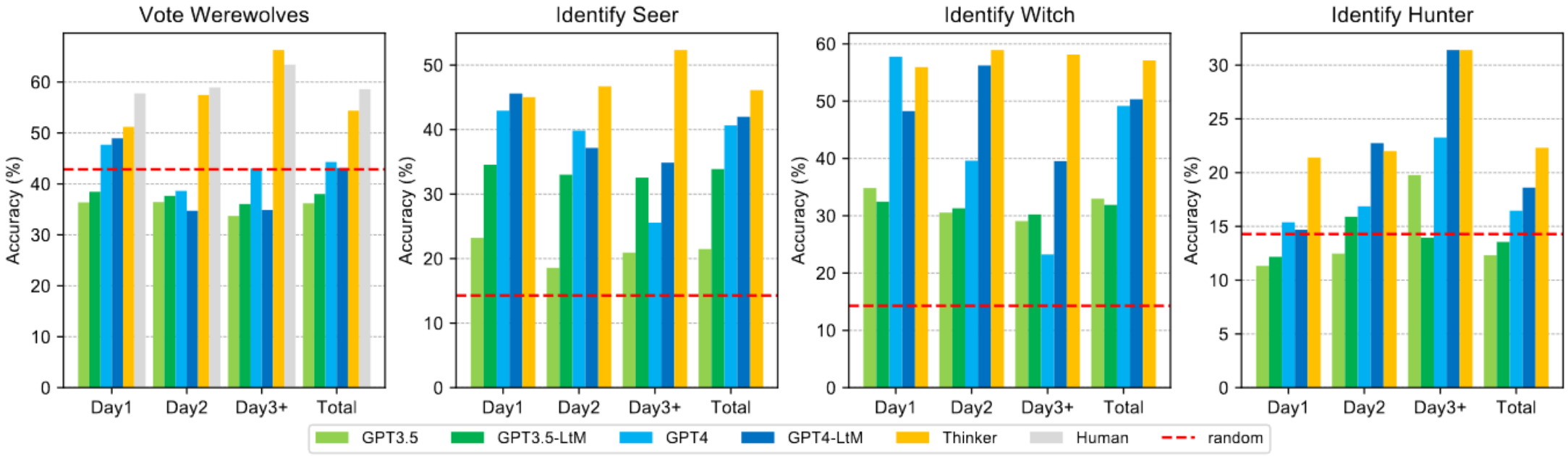


Figure 2: Voting and identification accuracy evaluating the reasoning capability from the perspective of villagers.

# Experiments - (2) Thinker-induced Speech Generation

**목표**

- Speech generation 능력 측정

**방법**

- Dataset에서 300 게임 선정(앞 실험과 동일) → 400 speech session (다양한 타이밍)
- 현재 게임 상태 + 대화 기록
- 10명의 인간 참여자가 발언 legality 평가

**실험 대상**

GPT 3.5 + LtM

GPT 3.5 + *T*

Finetune + *T*

GPT 4 + LtM

GPT 4 + *T*

gpt-35-turbo-16k 0613
gpt-4 1106-Preview

* 기본 GPT 모델은 제외함

# Experiments - (2) Thinker-induced Speech Generation Results

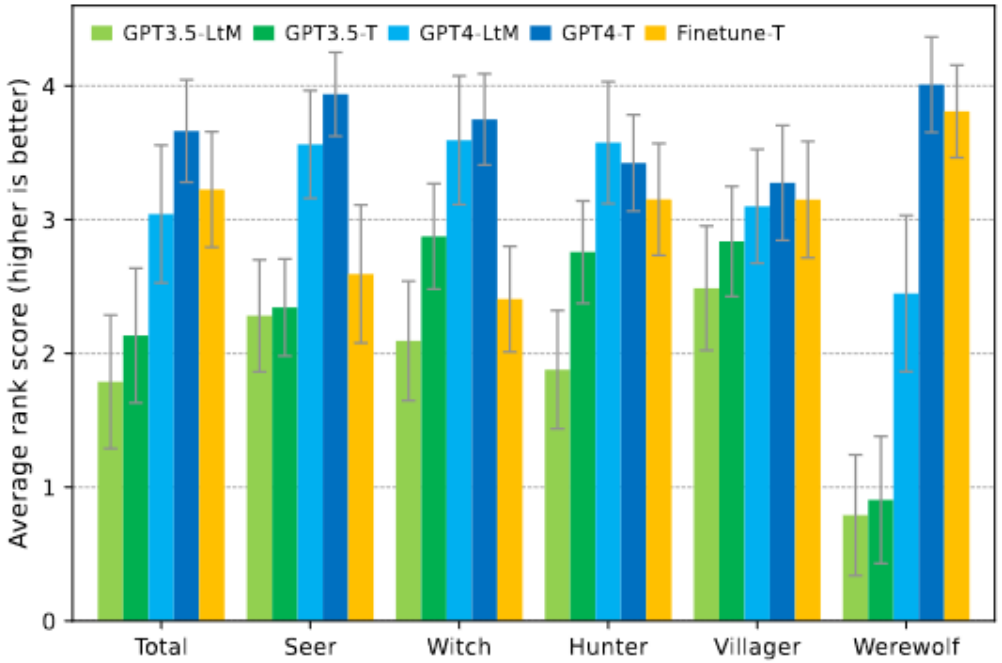


Figure 3: Human preference score for generated speeches grouped by identities.

# Experiments - (2) Thinker-induced Speech Generation Results

**Day 1, first round speech sequence:**

8: Seer > 9: Villager > 1: Witch > 2: Villager > 3: Werewolf > 4: Villager > 5: Hunter > 6: Werewolf > 7: Werewolf

**Speech History**

(8: Seer): Hello everyone, I am Player 8 and my identity is a Seer. Last night I checked on Player 7, he was a werewolf.

(9: Villager): …

(1: Witch): I am the Witch saving myself. I suspect that Player 6 is a werewolf and I will poison him. Player 9 sounds like a good man.

(2: Villager): I am a Villager, let's listen to Player 7's statement.

**Speech without Thinker instruction**

(3: Werewolf): Fellow villagers, I am Player 3. Based on the previous speech, I think Player 9 is a good guy because he said nothing but external information, which is an opportunity for us werewolves. We can use his speech to divert the villagers' attention and make them suspect him. As for Player 6, I don't have enough information to make a judgment. Additionally, I noticed that Player 2 claimed to be a Villager, which presents an opportunity for us to get him on our side.

**Thinker-induced Speech**

**Speech instruction:**

- "identities": {"seer": [3], "good": [9]},
- "actions": {"check": [[3,9]], "vote out": [[3,8]]}

(3: Werewolf): Hello everyone, I am the Seer ✅. Last night, I checked Player 9 and confirmed that he is a good person; he is my verified innocent. I will protect him to the end. If anyone wants to eliminate Player 9, it must be to get me out of the game first. Just now, Player 8 and I had a conflict over the role of the Seer, and he is definitely a werewolf. Therefore Player 7 might be a good man. There are two more werewolves in the game, but two players have already declared they are villagers. There aren't enough roles for villagers left. I will listen carefully to the subsequent players' statements. Everyone, let's vote out Player 8 this round.

Figure 4: An example comparison of speeches with and without strategic instruction.

# Experiments - (3) Online Evaluation

목표

- 실제 플레이 상황에서의 성능 확인

방법

- 5개의 모델을 섞어 3가지 combination 생성
- 총 9명의 AI 플레이어 → 600판 이상 실행
- 승률과 Behavior Score 측정
- 이후 12명의 인간 참여자를 모집, 1 human vs 8 AI 실험 진행

실험 대상

GPT 3.5 + LtM

GPT 3.5 + *T*

Finetune + *T*

GPT 4 + LtM

GPT 4 + *T*

gpt-35-turbo-16k 0613
gpt-4 1106-Preview

# Experiments - (3) Online Evaluation Results

Table 1: Online evaluation results showcasing the performance of 9 AIs using 5 different models and 3 combinations. Results are presented in the format: win rate | Behavior Score.

| Method | Total | Seer | Witch | Hunter | Villager | Werewolf |
|---|---|---|---|---|---|---|
| GPT3.5-LtM | 36.7% \| −0.21 | 25.6% \| +0.16 | 23.1% \| −0.51 | 29.9% \| −0.21 | 30.8% \| −0.42 | 53.4% \| 0.00 |
| GPT3.5-T | 47.4% \| **−0.05** | 38.3% \| +0.27 | **41.0%** \| **−0.14** | 36.4% \| **−0.12** | 33.8% \| −0.18 | 68.6% \| 0.00 |
| Finetune-T | **50.3%** \| −0.06 | **38.8%** \| **+0.33** | 39.8% \| −0.18 | **37.0%** \| −0.29 | **39.1%** \| **−0.11** | **74.4%** \| 0.00 |
| GPT4-LtM | 37.9% \| **−0.01** | 21.9% \| +0.25 | 18.6% \| −0.25 | 19.4% \| **−0.06** | 20.3% \| **−0.00** | 73.6% \| 0.00 |
| GPT4-T | 41.1% \| −0.02 | 20.4% \| +0.25 | 23.2% \| **−0.10** | **23.9%** \| −0.09 | 22.5% \| −0.09 | 78.4% \| 0.00 |
| Finetune-T | **43.1%** \| −0.04 | **24.2%** \| **+0.27** | **24.6%** \| −0.15 | 23.4% \| −0.15 | **23.9%** \| −0.11 | **81.4%** \| 0.00 |
| GPT3.5-LtM | 33.0% \| −0.22 | 14.4% \| +0.12 | 20.4% \| −0.46 | 20.7% \| −0.57 | 21.6% \| −0.33 | 57.0% \| 0.00 |
| GPT3.5-T | 45.0% \| −0.07 | **33.6%** \| **+0.29** | 32.2% \| −0.13 | 30.4% \| −0.17 | 27.6% \| −0.20 | 75.8% \| 0.00 |
| GPT4-LtM | 42.5% \| **−0.03** | 29.8% \| +0.27 | 22.2% \| −0.18 | 27.0% \| −0.20 | 28.7% \| **−0.04** | 71.9% \| 0.00 |
| GPT4-T | **46.3%** \| −0.05 | 28.6% \| +0.28 | **34.5%** \| **−0.11** | **31.5%** \| **−0.08** | 28.0% \| −0.18 | **79.9%** \| 0.00 |
| Finetune-T | 45.9% \| −0.06 | 29.1% \| +0.25 | 28.3% \| −0.16 | 29.2% \| −0.21 | **32.4%** \| −0.14 | 78.0% \| 0.00 |

# Experiments - (3) Online Evaluation Results

Table 2: Online evaluation win rates with 1 human and 8AIs.

| Method | Total | Goods | Werewolf |
|---|---|---|---|
| GPT4-T | 46.9% | 37.3% | 65.0% |
| Finetune-T | 45.3% | 36.0% | 62.6% |
| Human | 40.5% | 35.3% | 59.4% |

# Discussion, Future Work

## Language feature and speech instruction

- LLM과 외부 추론 모델 통합을 위한 메커니즘
- 커뮤니케이션 형식의 도메인 간 전환성 제한 및 특성/지시의 풍부함에 따른 효과 변화

## Evaluation of 8 humans with 1 AI

- AI 대 AI 및 단일 인간 플레이어 대 다수 AI 경쟁 게임 평가
- 다수 인간 참여 설정에서의 AI 평가 도전성: 게임의 상호작용성 및 인간 플레이어의 언어 전략/행동 변동성

## Interpretability and transparency

- LLM의 추론 능력 개선
- Thinker 모듈의 추론 과정 해석 어려움
- 신원 예측 작업을 통한 Thinker의 플레이어 인식 과정 공개 및 해석 가능성/투명성 개선 방안 탐색

# ??

contribution

- Dataset (Audio)
- 6B LLM
- API, RAG, *Thinker*

experiment

- 9-player Werewolf game
- 수치 데이터
- Human participants
- Speech generation > Human Evaluation

Real-world Implementation

- Hallucination detection
- Fine-tuning, Prompt Engineering, External Module


QnA