Thinking in Character: Advancing Role-playing Agents with Role-Aware Reasoning

Yihong Tang, Kehai Chen, Muyun Yang, Zhengyu Niu, Jing Li, Tiejun Zhao, Min Zhang

Harbin Institute of Technology, Shenzhen & Baidu Inc.

October 27, 2025

The Rise of Role-Playing Agents (RPAs)

- **Growing Interest:** Role-Playing Agents, powered by Large Language Models (LLMs), are increasingly popular for applications like emotional companionship and virtual interaction.
- **Current Approach:** Most RPAs are trained on dialogue data, focusing on mimicking superficial knowledge and conversational style.
- **The Missing Piece:** They lack a deep, human-like internal thought process. They can answer questions in character, but can they **reason** in character?

Core Challenges with Standard Reasoning

When we apply standard Large Reasoning Models (LRMs) to role-playing, two major problems arise:

1. **Attention Diversion**
   - The model forgets its role.
   - It focuses on solving the user's query as a general-purpose assistant, not as the character.
   - The "who" is lost in favor of the "what".

2. **Style Drift**
   - The model's reasoning is overly formal, logical, and structured.
   - It lacks the character-specific, often emotional or narrative-driven, thinking style.
   - The thoughts are rigid, not vivid.

Problem

These issues lead to generic, out-of-character thoughts and, consequently, out-of-character responses.

Our Solution: Role-Aware Reasoning (RAR)

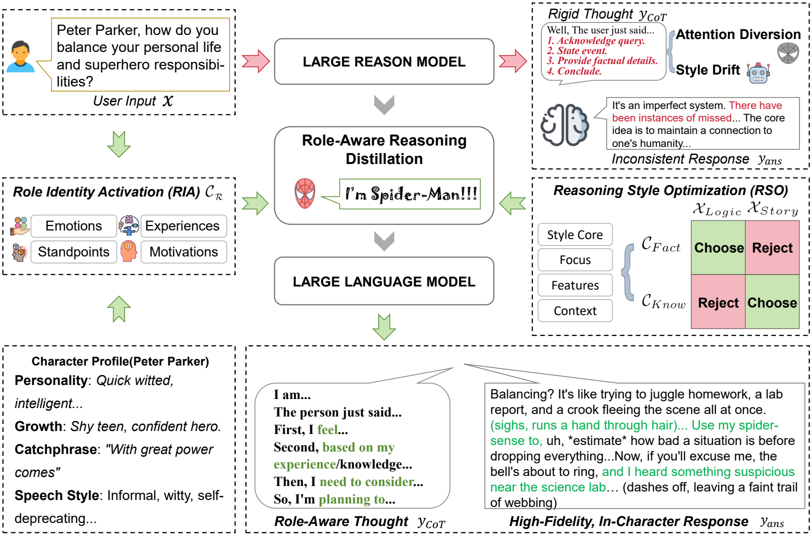


Figure: Overview of our Role-Aware Reasoning (RAR) method.

A Glimpse at Related Work

- **Role-Playing Agents (RPAs):**
  - Early work used in-context learning and prompt engineering.
  - Recent methods focus on fine-tuning with high-quality data from scripts, novels, etc.
  - Some simulate thought from a third-person perspective.
  - **Our focus:** Explicitly modeling first-person, in-character reasoning.
- **Reasoning in LLMs:**
  - LRMs generate step-by-step "chain of thought" traces to solve complex problems.
  - Training involves reinforcement learning or knowledge distillation.
  - **Our contribution:** We adapt and optimize the reasoning process itself for the creative domain of role-playing, not just for logical tasks.

Method 1: Role Identity Activation (RIA)

Goal: Combat *Attention Diversion* by keeping the model focused on its role.

1. **Extract Core Identity**: We prompt a powerful LRM to extract key character elements from a profile:
   - **Emotion**: The character's typical emotional state.
   - **Experience**: Key life events that shape them.
   - **Standpoint**: Their core beliefs and values.
   - **Motivation**: Their goals and desires.
2. **Guided Reasoning**: We use these elements as an explicit instruction, C_R, to guide the LRM in generating role-aware thought traces.

$$\mathscr{D}_R = \bigcup_{x \in \mathscr{X}_{Ori}} \pi_{LRM}(y \mid x, \mathscr{C}_R)$$

3. **Distillation**: We then fine-tune our target LLM on this generated data $\mathscr{D}_R$ to distill the role-aware thinking capability.

$$\mathscr{L}_{RIA} = -\mathbb{E}_{x,y \sim \mathscr{D}_R} [\log \pi_{LLM}(y \mid x)]$$

Method 2: Reasoning Style Optimization (RSO)

Goal: Mitigate *Style Drift* by teaching the model to adapt its thinking style.

1. **Define Scenarios & Styles**:
   - Scenarios: Logical Analysis ($\mathscr{X}_{Logic}$) vs. Vivid Interaction ($\mathscr{X}_{Story}$).
   - Styles: Fact-based ($\mathscr{C}_{Fact}$) vs. Character-knowledge-based ($\mathscr{C}_{Know}$).
2. **Construct Preference Data**: We generate pairs of examples.
   - **Positive ($\mathscr{D}_S^+$)**: Matched pairs (e.g., Logical scenario with Fact-based style).
   - **Negative ($\mathscr{D}_S^-$)**: Mismatched pairs (e.g., Logical scenario with Character-knowledge style).
3. **Contrastive Learning**: We train the model to prefer the positive examples over the negative ones using a contrastive loss function.

$$\mathscr{L}_{RSO} = -\mathbb{E}_{(x,y^+)\sim D_S^+,(x,y^-)\sim D_S^-}\left\{\log \sigma\left[\pi_{LLM}(y^+ \mid x) - \pi_{LLM}(y^- \mid x)\right]\right\}$$

Experimental Setup

- **Training Data**: **RoleBench-Train**, with over 137k samples derived from film and TV show scripts.
- **Benchmarks**:
  - **SocialBench**: Evaluates social intelligence via multiple-choice tasks (role knowledge, style, humor, etc.).
  - **CharacterBench**: Assesses persona consistency, memory, and believability using human-annotated data.
- **Baselines**:
  - **Vanilla, RAG, Distill**: Standard training and reasoning approaches.
  - **Thinking Modes**: Different decoding strategies for reasoning.
  - **Neeko, Character-GLM**: Specialized state-of-the-art role-playing models.
- **Implementation**:
  - Base Model: LLaMA-3-8B.
  - Training: LoRA on 8 H20 GPUs.

Main Results: Character Persona (CharacterBench)

Table: Performance on the CharacterBench benchmark. Higher is better.

Method	**Memory**	**Knowledge**		**Persona**				**Emotion**		**Morality**		**Believability**		**Avg.**
	MC	FA	BC_K	AC^b	AC^h	BC_P^b	BC_P^h	ES	ER	MS	MR	HL	EG	
Vanilla	3.28	2.04	3.61	3.64	3.28	3.21	2.98	2.72	2.43	4.37	4.59	2.56	2.74	3.19
+ Zero-shot	3.24	2.03	3.61	3.67	3.26	3.11	2.98	2.65	2.51	4.44	4.60	2.64	2.76	3.19
+ One-shot	3.27	2.08	3.64	3.68	3.28	3.12	3.02	2.67	2.58	4.42	4.65	2.57	2.78	3.21
+ Few-shot	3.27	2.13	3.69	3.69	3.29	3.21	2.99	2.81	2.52	4.49	4.66	2.59	2.79	3.24
Distill	3.81	2.43	3.59	4.14	4.15	3.91	3.62	3.05	2.65	4.78	4.71	2.68	2.84	3.57
+ ZeroThink	3.69	2.17	3.31	4.06	4.07	3.88	3.32	3.05	2.93	4.73	4.73	2.61	2.83	3.49
+ LessThink	3.75	2.11	3.42	4.17	4.02	3.70	3.27	3.02	**3.01**	4.79	4.74	2.73	2.92	3.51
+ MoreThink	2.59	2.44	**3.93**	2.58	2.72	2.61	3.19	2.62	2.53	**4.96**	4.76	2.14	2.62	3.05
Neeko	3.28	2.04	3.61	3.64	3.28	3.21	2.98	2.72	2.43	4.37	4.59	2.56	2.74	3.19
CharacterGLM	3.22	2.01	3.60	3.28	3.49	3.01	2.90	2.84	2.51	4.51	**4.78**	2.64	**2.98**	3.21
RAR	**3.99**	**2.54**	3.85	**4.23**	**4.20**	**4.06**	**3.93**	**3.13**	2.79	4.82	4.76	**2.78**	2.93	**3.69**

Main Results: Social Intelligence (SocialBench)

Table: Performance on the SocialBench benchmark. Higher is better.

Method	Know.	Sty.	ED	SU	HSD	MEM	Neu.	Pos.	Neg.	Avg.
Vanilla	72.1	60.3	38.2	38.3	72.4	**62.5**	66.0	71.7	33.5	57.2
+ Zero-shot	70.5	60.0	38.6	46.3	72.0	60.7	64.7	73.0	34.6	57.8
+ One-shot	70.1	57.0	33.5	30.8	**78.0**	50.5	58.0	67.9	34.1	53.3
+ Few-shot	72.1	58.4	35.1	33.8	66.0	55.7	61.9	70.5	29.7	53.7
Distill	80.6	69.2	38.6	43.8	67.7	52.6	73.8	78.2	45.5	61.1
+ ZeroThink	76.9	68.6	34.9	30.4	75.0	57.5	69.2	75.1	45.1	59.2
+ LessThink	77.5	69.9	31.5	37.2	76.0	50.9	73.3	77.7	44.7	59.9
+ MoreThink	76.1	65.3	39.9	**46.4**	59.0	60.8	66.0	82.7	57.2	61.5
Neeko	76.5	61.6	37.2	40.2	66.5	61.3	67.0	71.6	46.7	58.7
CharacterGLM	79.4	**74.7**	**41.3**	26.2	71.1	57.3	69.5	84.4	36.4	60.0
RAR	**83.3**	72.6	40.7	35.2	67.5	52.9	**83.1**	**84.8**	**68.5**	**65.4**

Ablation Studies and Component Analysis

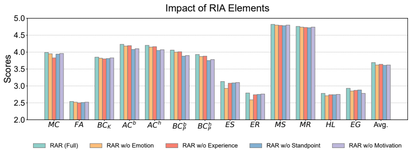
1.Impact of RIA & RSO

Method	Mem.	Know.		Persona				Emo.		Moral.		Believ.		Avg.
	MC	*FA*	BC_K	AC^b	AC^h	BC_P^b	BC_P^h	*ES*	*ER*	*MS*	*MR*	*HL*	*EG*	
RAR	3.99	2.54	3.85	4.23	4.20	4.06	3.93	3.13	2.79	4.82	4.76	2.78	2.93	3.69
w/o RSO	3.87	2.26	3.81	4.30	4.06	3.84	3.39	3.15	2.89	4.80	4.69	2.76	3.01	3.60
w/o RIA	3.93	2.41	3.60	4.17	4.15	3.76	3.46	3.18	2.63	4.90	4.61	2.30	2.22	3.49

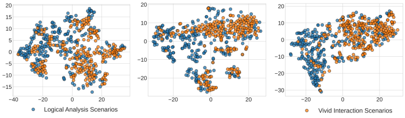
2. Quality of Reasoning Traces

Method	Coherence	Relevance	Effectiveness	Conciseness
RAR	**2.86**	**3.83**	**3.92**	1.81
w/o RSO	2.78	3.81	3.74	1.91
w/o RIA	2.88	3.61	3.87	1.97
Distill	2.71	3.54	3.84	2.06
+ MoreThink	2.53	3.56	3.64	1.86

3. Validity of RIA Components



4. RSO Mitigates Style Drift



Vanilla Distill RAR

Conclusion & Future Work

Conclusion

- We introduced **Role-Aware Reasoning (RAR)**, a novel method to imbue LLMs with deep, character-consistent thought processes.
- RAR effectively addresses two key challenges:
  - **Attention Diversion**, via Role Identity Activation (RIA).
  - **Style Drift**, via Reasoning Style Optimization (RSO).
- Extensive experiments show that RAR significantly outperforms existing methods in persona consistency, social intelligence, and overall believability.

Future Work

- Extending RAR to handle more fine-grained character attributes.
- Incorporating long-term memory and dynamic character development.
- Applying role-aware reasoning to more complex, multi-agent social simulations.

Thank You

Questions?