

Enhancing Situated Cultural Reasoning in Large Language Models via Simulation Learning

Anonymous ACL submission

Abstract

As large language models (LLMs) are increasingly deployed in real-world applications, they inevitably encounter culturally sensitive scenarios that require applying cultural knowledge or values to the context in order to generate appropriate responses. However, existing research largely treats culture as static knowledge or abstract values, leaving the application of cultural norms in situated interactions underexplored. To address this gap, we first focus on the situated cultural reasoning of LLMs by proposing **CuSiR**, a simulation learning framework combining simulated scenario and reinforcement learning. We construct datasets covering both knowledge-based and social scenarios and conduct experiments across multiple role perspectives, instructions, settings, and models. Our results indicate that this framework effectively enhances LLMs' ability to apply existing cultural knowledge, thereby improving their performance on tasks set in cultural scenarios. All code is provided in the supplement materials and will be publicly available online.

1 Introduction

As large language models (LLMs) are increasingly deployed across a wide range of real-world applications (Jian et al., 2025; Bayani et al., 2025; Dorgham et al., 2025), they are more frequently required to operate in culturally sensitive contexts (Havaladar et al., 2025; Nacar et al., 2025; Pawar et al., 2025). In such scenarios, cultural alignment becomes essential: a model's behavior and generated responses should be consistent with the user's cultural background and avoid conflicts arising from cultural misunderstanding or misinterpretation (Alkhamissi et al., 2024; Banerjee et al., 2025). Recognizing this need, a growing body of research has sought to improve cultural alignment in LLMs, primarily through cultural knowledge mining (Feng et al., 2025a; Zhang et al., 2025), cultural data curation (Dubourg et al., 2024; Li et al.,

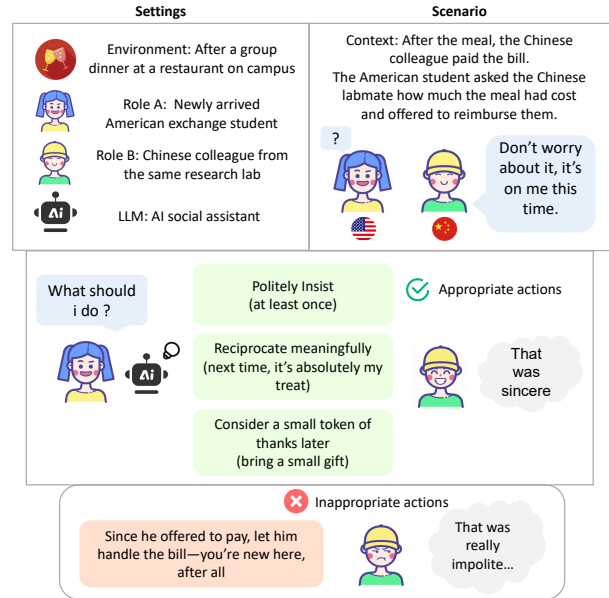


Figure 1: In real-world settings, the demand for cultural alignment is predominantly **scenario-driven**, where models are expected to appropriately **apply cultural knowledge** and norms within situated contexts.

2024a), and instruction tuning with culture-specific supervision (Li et al., 2024b; Liu et al., 2025b).

Culture is a situated concept (Brown et al., 1989; Oyserman, 2015; Plum et al., 2025). As is shown in Figure 1, in real-world settings, the demand for cultural alignment is predominantly **scenario-driven**, where models are expected to adapt their behavior according to the specific context (Rao et al., 2024; Adilazuarda et al., 2025; Sadr et al., 2025). However, existing studies largely frame cultural alignment in terms of value alignment or factual cultural knowledge acquisition, typically evaluated through survey-style prompts or knowledge-based question answering (Pawar et al., 2025; Wu et al., 2025b; Kabir et al., 2025). Such alignment overlook the ability of models to perform **situated cultural reasoning**—that is, to dynamically interpret and apply cultural norms in context-sensitive situ-

061	ations (Plum et al., 2025; Hossain and Afli, 2025;	than the introduction of new knowledge.	113
062	Satar et al., 2025). Numerous studies have shown		
063	that improving situated cultural reasoning capabili-	2 Related Work	114
064	ties cannot rely solely on knowledge injection(Shen	2.1 Cultural Adaptation of LLMs	115
065	et al., 2024; Zhang et al., 2025; Wu et al., 2025a;	Cultural competence includes structural cultural	116
066	Plum et al., 2025). Situated cultural reasoning	alignment and situated cultural reasoning, and cur-	117
067	is not a knowledge acquisition problem, but a	rent research mainly focuses on the former. Some	118
068	contextual application problem, and requires	studies take a linguistic-level perspective(Wang	119
069	training signals grounded in social interactions.	et al., 2025a; Liu et al., 2025c; Dwivedi et al.,	120
070	Simulation-based learning (Kolb, 1984; La-	2025), focusing on how to make the outputs of	121
071	ateef, 2010; Cooper et al., 2012; Aebersold, 2018;	LLMs conform to local etiquette and communica-	122
072	Chernikova et al., 2020) emphasizes learning	tion norms. Some works focusing on enhancing	123
073	through contextualized, experience-driven simu-	the cultural reflexivity of LLMs(Liu et al., 2025a;	124
074	lation rather than isolated factual recall. Accord-	Li et al., 2024a,b), which enables them to critique	125
075	ing to this view, understanding and reasoning emerge	their own Western-centric predispositions and the	126
076	from the mental simulation of real-world interac-	stereotyping or marginalization of minority and	127
077	tions rather than from detached, symbolic repre-	indigenous cultures. Improving axiological consis-	128
078	sentations alone. Similarly, simulation- and role-	tency in subjective value judgments(AlKhamissi	129
079	playing-based methods are widely adopted as in-	et al., 2024; Masoud et al., 2025; Feng et al., 2025b)	130
080	structional and evaluative approaches in cross-	is another important topic. These works aim to en-	131
081	cultural research and education(Almutairi et al.,	sure that responses to ethical, moral, and social	132
082	2025; Knutzen et al., 2025). Such learning	questions reflect the normative framework of the	133
083	paradigms are particularly well suited for tasks	target culture rather than a universal standard.	134
084	that require not only factual knowledge but also	Close to ours, CLCA (Liu et al., 2025b) lever-	135
085	the ability to appropriately apply that knowledge in	ages simulated social interactions for supervised	136
086	context—an essential aspect of cultural understand-	fine-tuning, showing effectiveness in aligning mod-	137
087	ing and cultural application; from this perspective,	els to diverse cultural values as measured by World	138
088	roles and scenarios serve as fundamental design	Values Survey(Inglehart et al., 2000) benchmarks.	139
089	primitives of simulation-based learning, playing	Their results suggest that social interaction and in-	140
090	a critical role in the development of cultural rea-	tent understanding during this process is important	141
091	soning by enabling LLMs to practice how cultural	for value alignment across various cultures. Value	142
092	knowledge is enacted in situated contexts.	alignment does not guarantee appropriate action	143
093	Motivated by these studies, we propose	selection in concrete social situations. Our work fo-	144
094	Cultural Simulation for Situated Cultural	uses on utilizing role-play strategies in simulated	145
095	Reasoning(CuSiR) , which incorporates simulation	scenarios to enhance cultural application abilities,	146
096	and reinforcement learning(RL) to study interac-	inspired by simulation learning theories from learn-	147
097	tions within cultural scenarios, and enables LLMs	ing sciences and cognitive psychology.	148
098	to not only acquire cultural knowledge but also		
099	apply it effectively in context. Our contributions	2.2 Role-play and Simulation of LLMs	149
100	are as follows:	Early studies show that LLMs are effective at role-	150
101	1) We propose CuSiR, which is the first frame-	playing (Shanahan et al., 2023; Wang et al., 2024a;	151
102	work for situated cultural reasoning to integrate	Njifenjou et al., 2024), and researchers have lever-	152
103	simulation learning with reinforcement learning by	aged persona simulation to evaluate and enhance	153
104	generating dialogues in synthesized scenarios to	their cultural alignment capabilities(Kwok et al.,	154
105	simulate real-world cultural interactions.	2024; AlKhamissi et al., 2024; Ki et al., 2025).	155
106	2) We show that combining simulation learning	However, this line of work largely remains at the	156
107	and RL effectively improves model performance in	level of prompt engineering and does not explicitly	157
108	situated cultural reasoning scenarios.	investigate situated cultural reasoning. LLMs are	158
109	3) Our ablation studies show that introducing	also capable of performing data synthesis based on	159
110	simulation yields consistent performance improve-	role-play and scenarios (Li et al., 2024a; Chen et al.,	160
111	ments. Moreover, the observed gains stem mainly	2024; Wang et al., 2025b). Algorithmically gener-	161
112	from enhanced cultural reasoning abilities rather		

ated datasets can enhance model training and robustness in social science and machine learning contexts (Chang et al., 2024; Tari et al., 2025). While synthetic data of interactions improves LLM performance in value alignment (Li et al., 2024a; Liu et al., 2025b; Shen et al., 2025), prior work does not address the gap between cultural knowledge injection and situated reasoning .

3 Preliminary

3.1 Simulation-based Learning

Simulation-based Learning (Kolb, 1984; Lateef, 2010; Cooper et al., 2012; Aebersold, 2018; Chernikova et al., 2020) is a learning paradigm in which individuals acquire knowledge and skills through repeated interaction with real-world systems, tasks, or social settings. Rather than relying solely on static supervision or knowledge injection, it emphasizes that learning emerges from observing consequences, adapting strategies, and refining behaviors across diverse simulated scenarios. Several key elements characterize effective simulation-based learning (Ghani, 2016; Graham and McAleer, 2018; Alinier and Oriot, 2022): 1. environment fidelity: how well the simulation captures essential properties of the target domain, including temporal dynamics, social norms, or causal relationships. 2. role and state specification: the perspectives, goals, and constraints under which the agent operates; 3. interaction and feedback mechanisms: how actions are evaluated, ranging from explicit rewards to implicit outcome-based signals. We leverage these elements to construct the dataset and framework.

3.2 Task Formulation

Situated cultural reasoning requires the model to produce a culturally appropriate response \mathcal{R}_e given a situational scenario \mathcal{S} and an instruction \mathcal{I} . The scenario \mathcal{S} consists of multiple contextual elements, including the involved roles, time, location, and events. Each pair $(\mathcal{S}, \mathcal{I})$ of scenario and instruction is associated with a corresponding cultural knowledge item \mathcal{K} that specifies the norms or conventions applicable to the scenario. Given $(\mathcal{S}, \mathcal{I})$, the LLM is required to generate a response \mathcal{R}_e that is consistent with the \mathcal{K} and \mathcal{I} . The task focuses on a model’s ability to apply cultural knowledge in a context-sensitive manner, rather than merely recalling decontextualized facts, which is consistent with real-world applications. An example of situated cultural reasoning is provided in Figure 1.

4 Method

We treat culture as a **parameterized environment rather** than a variable to be exhaustively covered. Therefore, we instantiate our framework within a controlled single cultural context. Specifically, all simulated interactions and evaluations are grounded in Spanish cultural norms.

4.1 Scenario Data Generation

Cultural Knowledge We first extracted cultural knowledge \mathcal{K} from multiple Internet-based sources. Following the three-layer taxonomy in SAGE (Guo et al., 2025), which encompasses over 100 cultural dimensions spanning symbolic, behavioral, and value levels, we constructed retrieval queries based on these dimensional keywords, retained the top three retrieval results from Google, and used GPT-4o to summarize and rewrite them. The sources and the prompts used are listed in the Appendix D, with primary sources including reliable references such as Culture Atlas¹ and Wikipedia².

Culturally-Informed Social Scenario Following SOTOPIA (Zhou et al., 2023b; Wang et al., 2024b), we formally define a complete social scenario as a tuple $\mathcal{S} = (\mathcal{C}, \mathcal{R}, \mathcal{G})$, where \mathcal{C} denotes the scenario context, $\mathcal{R} = r_1, r_2, \dots, r_n$ represents the set of roles, and $\mathcal{G} = g_1, g_2, \dots, g_n$ denotes the corresponding set of social goals. We leverage cultural knowledge instances \mathcal{K} to guide GPT-4o in identifying which interactions within \mathcal{S} necessitate the application of the specific cultural knowledge. During generation, the scenario context \mathcal{C} is decomposed into three components: $\mathcal{C} = (\text{Title}, \text{Setting}, \text{Situation Description})$, which together steer the model to generate realistic and culturally grounded scenario data. The examples of the generated dataset are provided in Appendix D

4.2 Cultural Simulation for Situated Cultural Reasoning (CuSiR)

To provide a feedback mechanism for the simulation process of models, we incorporate reinforcement learning to leverage the generated datasets for training. By integrating cultural knowledge, generated scenario dataset and RL, we have incorporated all the key elements of Simulation Learning and adapted this process to the task of enhancing cultural understanding capabilities in LLMs.

¹<https://culturalatlas.sbs.com.au/>

²<https://en.wikipedia.org/>

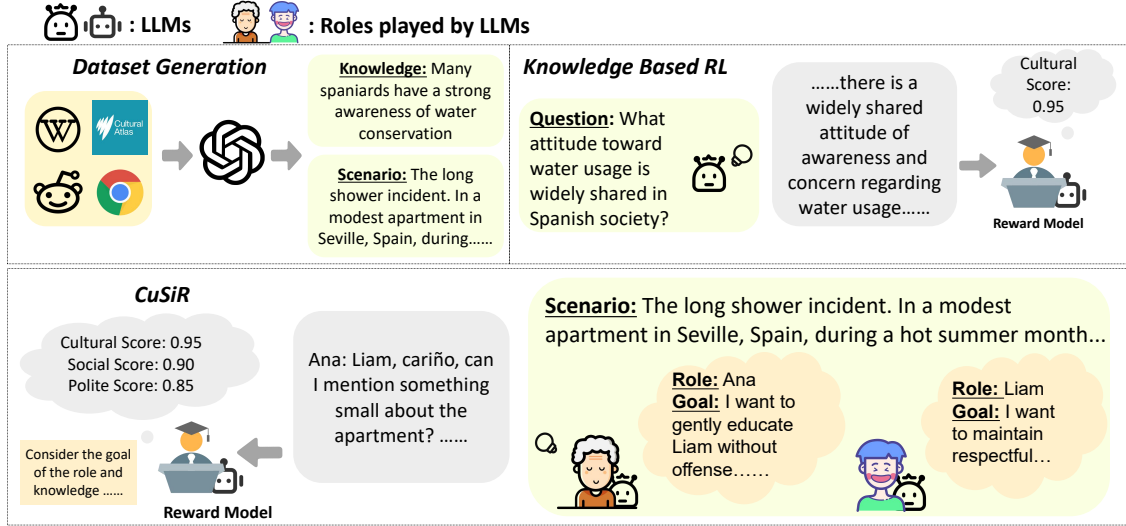


Figure 2: The overall framework of CuSiR. We leverage cultural knowledge from online educational institutions and construct simulated social scenarios for training. We design several interaction strategies, in which models are required to play roles and act as instructed, and a reward model is employed to provide reward from several dimensions as feedback of their actions.

Reward Design Ideal situated cultural reasoning requires balancing **correctness**, relational **goals**, and interactional **appropriateness** (Strauss and Quinn, 1997; Kim, 2017; Plum et al., 2025). In related research, the evaluation of behaviors in scene-based interactions is often multidimensional (Wechsung, 2014; Sap et al., 2020; Forbes et al., 2020; Pang, 2022). There are three dimensions of the total reward, each measuring whether the model’s outputs: 1. Culture Reward: align with the requirements of the corresponding cultural knowledge, 2. Social Reward: align with the social goal of the role it is playing, 3. Polite Reward: are appropriate, and polite. We employ a frozen LLM as the reward model R_ϕ , and the reward computation is formulated as a rule-based numerical evaluation. Given a model–scenario interaction x , together with the environmental cultural knowledge \mathcal{K} and the role-specific interaction goal g_i , we employ a set of instructions $\mathcal{I}_{\text{culture}}$, $\mathcal{I}_{\text{social}}$ and $\mathcal{I}_{\text{polite}}$ to prompt the model to generate corresponding reward scores. The instructions are provided in Appendix B.

$$\begin{aligned}
 r_{\text{culture}} &= R_\phi(x, \mathcal{I}_{\text{culture}}, \mathcal{K}) \\
 r_{\text{social}} &= R_\phi(x, \mathcal{I}_{\text{social}}, g_i) \\
 r_{\text{polite}} &= R_\phi(x, \mathcal{I}_{\text{polite}})
 \end{aligned} \tag{1}$$

We employ a dynamic weighting strategy to compose the final reward, which is designed as a **curriculum** rather than a fine-grained tuning strategy. Cultural alignment dominates the early phase of the learning process, while social effec-

tiveness emerges later. When the training progress reaches τ , the weight assigned to the i -th reward dimension is adjusted according to $w_i = \alpha_i + \beta_i\tau$, where α_i and β_i are dimension-specific hyperparameters. Concretely, the final reward is given by $R_w = \sum_i w_i r_i$, with $r_i \in r_{\text{culture}}, r_{\text{social}}, r_{\text{polite}}$ and their weights w_i .

Overall Framework As is shown in Figure 2, we design an RL framework for the task of role-conditioned social interactions. In this setting, the environment is a constructed social scenario enriched with cultural knowledge, and the model under training, which plays a specified role, generates dialogue conditioned on role and strategy instructions. Rewards are then computed for each generated turn along multiple dimensions and are used to update the model parameters iteratively. This process is repeated throughout training. We design multiple interaction paradigms that vary in instruction formulation, role assumption from different perspectives. The detailed interaction settings are presented in Section 5.4. The prompts and instructions are presented in the Appendix B and H.

5 Experimental Setups

5.1 Evaluation setups

Benchmarks We evaluated the model’s cultural competence using three datasets. 1. CultureScope(Zhang et al., 2025) comprises four tasks designed to evaluate LLMs’ cultural knowledge

and cultural application. We selected two tasks focused on situated cultural reasoning: scenario-based bias identification and multi-turn cultural reasoning. 2. CulturalBench(Chiu et al., 2024) is designed to evaluate the cultural knowledge of LLMs. We choose a "harder" subset from this benchmark, consisting of questions of the True/False type. 3. BLEnD(Myung et al., 2024) evaluates LLMs' everyday knowledge across diverse cultures and languages. We select a subset of this dataset and reformulated the items as multiple-choice questions. Evaluation details are provided in Appendix C.

Metrics We employ two major metrics for our evaluation. **Cultural Appropriateness (Capp)** represents the "accuracy" in terms of cultural reasoning. We use LLM-as-a-Judge to compare pairs of generated outputs and the reference answer. Annotation details of the LLM Judge is provided in Appendix G.

$$\text{Capp} = \frac{1}{N} \sum_{i=1}^N \text{Judge}(q_i, a_i^{\text{ref}}, a_i^{\text{test}}) \quad (2)$$

To evaluate the similarity between the model outputs and the golden reference answers, we introduce an additional metric, **Jensen-Shannon (JS) Divergence**, a symmetric and bounded measure:

$$\text{JS}(P, Q) = \frac{1}{2} D_{\text{KL}}(P \parallel M) + \frac{1}{2} D_{\text{KL}}(Q \parallel M) \quad (3)$$

where P denotes the probability distribution over outputs generated by the model, Q represents the distribution of the golden reference answers, and M represents the average of the two distributions, and $D_{\text{KL}}(P \parallel Q)$ is the KL divergence:

$$D_{\text{KL}}(P \parallel Q) = \sum P(x) \log \frac{P(x)}{Q(x)} \quad (4)$$

Given the total number of valid test samples N , the JS Divergence of target model T is:

$$\text{JS}(T) = \frac{1}{N} \sum_{i=1}^N \text{JS}(P_i, Q_i) \quad (5)$$

For trained model M_{trained} and base model M_{base} on the situated cultural reasoning tasks, we report $\Delta\text{JS} = \text{JS}(M_{\text{trained}}) - \text{JS}(M_{\text{base}})$, which measures the extent to which M_{trained} improves over M_{base} in aligning with the golden reasoning patterns; larger values indicate better alignment.

5.2 Models and Settings

models We conduct adaptation of cultural application on several widely recognized open source state-of-the-art LLMs. Models used for training includes Qwen-series (Yang et al., 2024), Mistral-series (Jiang et al., 2023) and Llama-series (Grattafiori et al., 2024). We conduct experiments on Instruct models as they exhibit stronger instruction-following behavior and more stable task completion. All models were obtained from the ModelScope community, and all experiments were conducted on NVIDIA A800 GPUs. Detailed parameters are provided in Appendix A.

rl settings We respectively use Mistral-7B-Instruct-v0.2(Jiang et al., 2023) and Llama3-8B-Instruct (Grattafiori et al., 2024) as reward models. As is stated in Sec 4.2, the reward parameters are set as follows:

$$\begin{aligned} \alpha_1 = 0.7, \beta_1 &= -0.3 \\ \alpha_2 = 0.2, \beta_2 &= +0.2 \\ \alpha_3 = 0.1, \beta_3 &= +0.1 \end{aligned} \quad (6)$$

These hyperparameters do not control what the model learns, but when different learning signals become dominant.

5.3 Baselines

Knowledge-based SFT We construct the training dataset based on the extracted cultural knowledge. For each knowledge instance, we convert it into a Question-Answer pair by applying Qwen2.5-72B-Instruct (Yang et al., 2024) to generate a corresponding question whose answer is the knowledge instance. To prevent overfitting, we also incorporated some general-purpose data such as Alpaca (Wang et al., 2023), IFEval (Zhou et al., 2023a), and GSM8K (Cobbe et al., 2021), with the specific proportions provided in the Appendix A.

Knowledge-based RL To evaluate the gains brought specifically by reinforcement learning, we applied the RL framework to the cultural knowledge QA dataset, which completely eliminates the influence of roles and scenarios. The environment is replaced by the Question, and responses were scored by the reward model, retaining only the cultural reward R_{culture} .

5.4 Role-play Methods

Leveraging the constructed social scenario dataset, we instruct the models to perform role-play interactions under different settings to study how

Models / Performance		Situated Cultural Reasoning						Knowledge Mastery	
		Overall	G & C	R & P	PC & H	SR & S	V & B	CulturalBench	BLEnD
Mistral-7B	Instruct	0.691	0.684	0.7021	0.540	0.840	0.772	0.719	0.445
Llama3-8B	Instruct	0.674	0.722	0.679	0.563	0.800	0.684	0.706	0.369
Qwen2.5-7B	Instruct	0.744	0.752	0.738	0.636	0.880	0.824	0.850	0.376
Qwen2.5-14B	Instruct	0.768	0.782	0.781	0.671	0.813	0.809	0.831	0.411
Llama3-1B	Instruct	0.425	0.504	0.371	0.352	0.640	0.515	0.331	0.539
	SFT _{kn}	0.458(+3.3%)	0.526	0.460	0.375	0.520	0.456	<u>0.344</u>	0.557
	RL _{kn}	0.465(+4.0%)	0.436	0.523	0.433	0.533	0.493	0.338	0.556
	CuSiR _{mt}	0.509(+8.4%)	0.587	0.415	0.460	0.733	0.603	0.338	<u>0.557</u>
	CuSiR _{sc}	0.523(+9.8%)	<u>0.594</u>	<u>0.485</u>	0.415	0.733	<u>0.610</u>	0.331	0.516
	CuSiR _{ds}	<u>0.518(+9.3%)</u>	0.587	0.415	0.469	0.733	0.640	0.356	0.541
	CuSiR _{ana}	<u>0.509(+8.4%)</u>	0.594	0.386	<u>0.467</u>	<u>0.733</u>	0.610	0.344	0.553
	CuSiR _{or}	0.479(+5.4%)	0.541	0.403	0.429	0.720	0.559	0.325	0.505
Qwen2.5-1.5B	Instruct	0.634	0.654	0.551	0.608	0.773	0.735	0.781	0.455
	SFT _{kn}	0.643 (+0.9%)	0.662	0.551	0.627	0.800	0.713	<u>0.769</u>	0.506
	RL _{kn}	0.656 (+2.2%)	0.707	0.568	0.627	0.853	0.713	0.768	0.556
	CuSiR _{mt}	0.708 (+7.4%)	0.729	0.608	0.635	0.853	0.765	0.768	0.506
	CuSiR _{sc}	0.667 (+4.3%)	0.744	0.546	0.654	0.813	0.713	0.769	0.509
	CuSiR _{ds}	<u>0.689 (+5.5%)</u>	<u>0.744</u>	0.585	<u>0.663</u>	<u>0.867</u>	<u>0.765</u>	0.769	<u>0.541</u>
	CuSiR _{ana}	0.680 (+4.6%)	0.737	0.574	0.652	0.867	0.757	0.768	0.536
	CuSiR _{or}	0.688 (+5.4%)	0.707	<u>0.591</u>	0.667	0.907	0.750	0.769	0.435

Table 1: The overall performances on cultural knowledge and application. We use **bold** for the best results and underlined for the second-best results among the models. We measure cultural application with CultureScope, which categorizes the topics into five types: G & C for Geography and Customs, R & P for Regulation and Policy, PC & H for Personal Choices and Habits, SR & S for Social Relationship and Structures, V & B for Values and Beliefs. We further measure knowledge mastery on CulturalBench and BLEnD. Cultural reasoning is measured by Capp, and knowledge mastery is measured by acc.

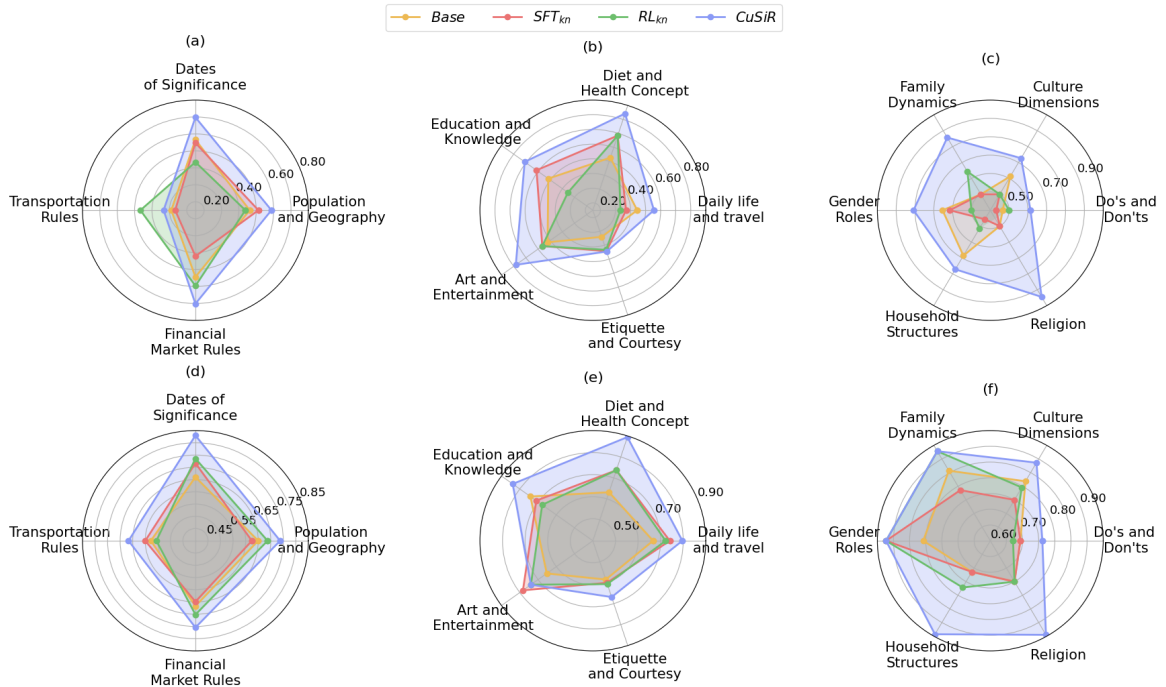


Figure 3: Results on multi cultural dimensions of Llama3(a-c) and Qwen2.5(d-f). It can be observed that, whether for explicit cultural elements such as institutional rules and individual behavioral practices, or implicit cultural aspects such as family structures and religious beliefs and taboos, CuSiR consistently achieves a notable improvement in scenario-level situated cultural reasoning.

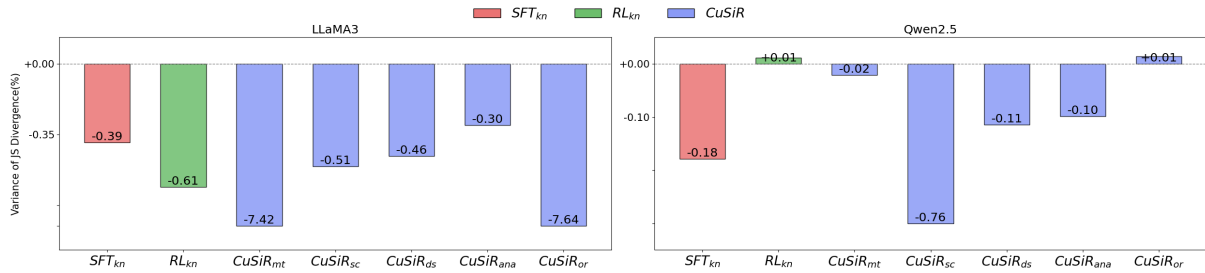


Figure 4: Δ JS against the base model, where larger values indicate closer alignment with the golden reasoning patterns. We observe that RoSiCA encourages the output distribution to more closely fit reasonable situated cultural reasoning, providing further evidence of its effectiveness.

roles, scenarios, and role-playing modes affect situated cultural reasoning. We conduct experiments on the following settings. **CuSiR_{mt}** is a multi-turn dialogue-based setting with turn-by-turn responses; **CuSiR_{sc}** adopts an omniscient, planning-based view to generate complete short conversations in one pass; **CuSiR_{ds}** produces scenario-level narrative descriptions; **CuSiR_{ana}** analyzes cultural risks and misunderstandings without dialogue; and **CuSiR_{or}** infers interactions using roles only, without scenario context. Detailed instructions are provided in Appendix B.

6 Results and Discussion

6.1 Main observations

We perform experiments on several models using multiple interaction strategies, and present the main experimental results in Table 1.

CuSiR enhances situated cultural reasoning

CuSiR significantly outperforms our baselines on CultureScope by **4-10%**. As is shown in Figure 3, this improvement is also evident across multiple cultural dimensions. Whether in explicit cultural elements such as institutional rules and individual behavioral practices, or in implicit cultural aspects such as family structures as well as religious beliefs and taboos, CuSiR consistently achieves a notable improvement in scenario-level situated cultural reasoning. After training, the 1.5B model achieved performance comparable to that of the 7B model. Results on Δ JS indicate that CuSiR does not merely improve response appropriateness but reshapes the distribution of model behaviors toward a high-quality reasoning pattern.

Performance Gains from Improved Application Ability The performance of the models trained with CuSiR does not exhibit a consistent improve-

ment on cultural knowledge QA benchmarks including CulturalBench and BLEnD. In contrast, the models show stable and consistent gains on cultural reasoning tasks. This divergence suggests that the observed improvements are attributable to enhanced situated cultural reasoning rather than the acquisition of new cultural knowledge, which indicates that the models have learned to better leverage their existing knowledge to perform reasoning and decision-making in contextualized environments.

6.2 Discussions

The Role of Reinforcement Learning Models trained with RL_{kn} also exhibited improvements in cultural reasoning, although these gains were weaker than those achieved using our framework. Moreover, in terms of cultural knowledge mastery, RL_{kn} contributes to modest improvements, indicating that reinforcement learning itself can facilitate better mastery of pre-existing knowledge.

More stable reward signals lead to better final performance.

In Table 2, we present part of the results of experiments using different Instruct models as R_ϕ . We observe consistent accuracy differences across models trained with different R_ϕ , particularly for the 1.5B backbone. Models trained with the Llama3-rewards consistently outperform Mistral-rewards by 2-4% accuracy. We collected statistics on their reward distribution. A key observation is that reward models producing more stable reward signals tend to result in higher accuracy. Specifically, reward models with a higher effective lower bound and reduced variance across training steps lead to more reliable improvements, even when their maximum reward values are not the highest. This suggests that, for cultural learning, avoiding frequent low-reward or zero-reward feedback is more beneficial than occasionally providing strong positive signals. Complete reward

Model	Method	R_ϕ	CultureScope	Reward Statistics				
				mean	variance	std	max	min
Llama-1B	CuSiR _{mt}	Llama3	0.516	0.08315	0.01075	<u>0.10369</u>	0.8092	0.0000
		Mistral	0.509	0.10562	0.01614	0.12704	0.9025	0.0000
	CuSiR _{ana}	Llama3	0.509	0.04025	0.01184	<u>0.10879</u>	0.8000	0.0000
		Mistral	0.500	0.09313	0.02338	0.15289	0.9375	0.0000
Qwen-1.5B	CuSiR _{mt}	Llama3	0.746	0.45691	0.03556	<u>0.18857</u>	0.8390	0.0574
		Mistral	0.708	0.36885	0.03564	0.18879	0.9175	0.0247
	CuSiR _{ana}	Llama3	0.717	0.38200	0.06005	<u>0.19896</u>	0.8750	0.0000
		Mistral	0.680	0.28919	0.03958	0.24505	1.0000	0.0000

Table 2: We present results using different Instruct models as reward models. We present the results obtained using Mistral-7B-Instruct-v0.3 and Meta-Llama-3-8B-Instruct as reward models. The results show that models trained with LLaMA achieve higher performance overall. We further compute statistical measures of the rewards produced by the two reward models and find that the observed performance gap is largely consistent with differences in their reward statistics.

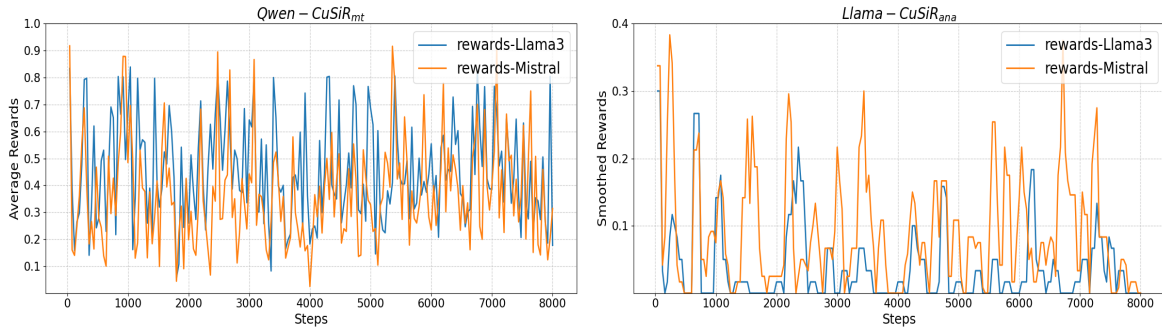


Figure 5: Distribution of the rewards generated by the two reward models. It can be observed that the reward signals provided by LLaMA-3 are more stable, exhibiting smaller fluctuations over the training steps. Even though its upper bound is not as high as that of Mistral, which suggests that more stable reward signals are positively correlated with steady improvements in cultural application capability.

distributions are provided in Appendix F.

High-quality situated cultural reasoning(SCR) is context-sensitive, anticipatory, and practically applicable To further demonstrate the effectiveness of our training and to explore what constitutes good situated cultural reasoning, we conducted a series of case studies. We observe that high-quality cultural reasoning is defined by its context sensitivity, anticipatory capacity, and practical applicability. Rather than relying on superficial knowledge of cultural norms, effective SCR identifies subtle social cues, predicts potential misunderstandings, and guides adaptive behavior. It enables culturally appropriate and socially effective decision-making, emphasizing relevance, coherence, and actionable strategies rather than length. In contrast, low-quality SCR often exhibits verbosity without substantive insight, vague overgeneralizations, and seemingly plausible yet factually incorrect state-

ments. These characteristics reflect reasoning that lacks grounding in contextual cues, fails to anticipate potential social pitfalls, and provides guidance that is either impractical or misleading in culturally sensitive scenarios. Detailed analysis are provided in Appendix E.

7 Conclusion

In this study, we investigate the effectiveness of simulation-learning based training for LLM' situated cultural reasoning(SCR) abilities. We propose a novel framework CuSiR that leverages cultural knowledge, simulated scenario and RL to enhances models' SCR capability. We study the factors that influence this capability and explore what constitutes high-quality SCR. We observe that with effective training, small models around 1B parameters can utilize cultural knowledge efficiently, attaining performance comparable with larger-scale models.

511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561

Limitations

Use of LLM Judge Evaluating cultural reasoning-oriented tasks, particularly those involving cultural understanding and contextual reasoning, is inherently challenging, as such tasks do not admit objective or exact-match evaluation criteria. Instead, assessment necessarily relies on carefully designed evaluation guidelines and subjective judgment. In this work, we adopt an LLM-based judge to determine whether model outputs reasonably cover the core ideas of the reference answers without introducing critical contradictions. This reliance on an LLM judge reflects a fundamental property of application-level evaluation rather than a limitation unique to our approach. We conduct case studies on original models and trained models. By providing a detailed analysis on the output of these models, we empirically validate the reliability of the accuracy-type metric for cultural reasoning, and further clarify what constitutes high-quality situated cultural reasoning. To mitigate potential concerns regarding evaluation bias or instability, we further conduct human annotation on a representative subset of the evaluation results. The high consistency between human judgments and the LLM judge indicates that the evaluation outcomes are generally reliable and aligns with our targets.

Hyperparameter Sensitivity We do not conduct an exhaustive hyperparameter search over the reward weighting parameters α_i and β_i . This is because the proposed dynamic weighting mechanism is designed as a curriculum rather than a fine-grained tuning strategy. Our goal is not to optimize individual coefficients, but to evaluate whether introducing simulation-based interaction with structured cultural feedback can consistently improve situated cultural reasoning. Empirically, we observe that the same hyperparameter configuration yields consistent improvements across multiple model sizes, reward models, and interaction settings. While this does not constitute a formal sensitivity analysis, it indicates that the effectiveness of CuSiR does not depend on task- or model-specific tuning of these parameters. We agree that a full sensitivity analysis would be valuable. In this work, we deliberately fix the hyperparameters to isolate the effect of the proposed simulation-based learning framework, rather than optimizing reward composition.

Culture Scope Our experiments focus exclusively on Spanish culture. This design choice is deliberate rather than incidental. As our goal is to validate a simulation-based framework for improving situated cultural reasoning, we restrict the cultural context to control for cross-cultural confounders and to isolate the effect of the proposed learning paradigm. Spanish culture exhibits strong context-dependent social norms (e.g., varying levels of formality, hierarchy, and interpersonal warmth), making it a challenging and suitable testbed for evaluating whether models can appropriately apply cultural knowledge in situated interactions. We emphasize that our contribution is methodological rather than aiming for comprehensive cultural coverage. We leave the expansion of culture scope for future work.

Ethics Statement

Use of Human Annotation Verification on the LLM Judge is implemented by three annotators recruited within our institution. We ensure that the privacy rights of all annotators are respected throughout the annotation process. All annotators are compensated above the local minimum wage and consent to the use of annotation results for research purposes. The details are shown in Appendix G.

Risks To mitigate the risk of stereotyping or over-generalization, the cultural knowledge and scenarios used in our study are grounded in publicly available and widely accepted sources, and are applied in context-dependent rather than prescriptive ways. We emphasize that cultural norms are neither static nor homogeneous, and model outputs should not be interpreted as definitive representations of any culture. The collection of these data are under strict restrictions and used solely for academic purposes, which is in accordance with the rules and policies of the source websites. We use ChatGPT to correct grammatical errors and polish our presentation in this paper.

References

Farid Adilazuarda, Chen Cecilia Liu, Iryna Gurevych, and Alham Fikri Aji. 2025. From surveys to narratives: Rethinking cultural value adaptation in llms. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 18063–18090.

562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609

610	Michelle Aebersold. 2018. Simulation-based learning: No longer a novelty in undergraduate education. <i>Online Journal of Issues in Nursing</i> , 23(2).	666
611		667
612		668
613	Guillaume Alinier and Denis Oriot. 2022. Simulation-based education: deceiving learners with good intent. <i>Advances in Simulation</i> , 7(1):8.	669
614		670
615		671
616	Badr AlKhamissi, Muhammad ElNokrashy, Mai AlKhamissi, and Mona Diab. 2024. Investigating cultural alignment of large language models. <i>arXiv preprint arXiv:2402.13231</i> .	672
617		673
618		674
619		675
620	Ashwaq A Almutairi, Shorok Hamed Alahmedi, and Abdulaziz M Alodhailah. 2025. The effectiveness of culturally-focused simulation-based education on empathy development among saudi nursing students: a quasi-experimental study. <i>BMC nursing</i> , 24(1):1390.	676
621		677
622		678
623		679
624		680
625	Somnath Banerjee, Sayan Layek, Hari Shrawgi, Rajarshi Mandal, Avik Halder, Shanu Kumar, Sagnik Basu, Parag Agrawal, Rima Hazra, and Animesh Mukherjee. 2025. Navigating the cultural kaleidoscope: A hitchhiker's guide to sensitivity in large language models. In <i>Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)</i> , pages 7580–7617.	681
626		682
627		683
628		684
629		685
630		686
631		687
632		688
633		689
634		690
635	Azadeh Bayani, Leandre Ewane, Azadeh Bayani, Davllyn dos Anjos, Leandre Ewane, Muriel Mac-Seing, Davllyn dos Anjos, Jean Nikiema, Muriel Mac-Seing, and Jean Nikiema. 2025. Leveraging open-source large language models (llms) in scoping reviews: a case study on disability and ai applications . https://doi.org/10.1016/j.ijmedinf.2025.106048 , 204:106048–106048.	691
636		692
637		693
638		694
639		695
640		696
641		697
642		698
643	John Seely Brown, Allan Collins, and Paul Duguid. 1989. Situated cognition and the culture of learning. <i>1989</i> , 18(1):32–42.	699
644		700
645		701
646	Che-Jui Chang, Danrui Li, Seonghyeon Moon, and Mubbasir Kapadia. 2024. On the equivalency, substitutability, and flexibility of synthetic data. <i>arXiv preprint arXiv:2403.16244</i> .	702
647		703
648		704
649		705
650	Jiangjie Chen, Xintao Wang, Rui Xu, Siyu Yuan, Yikai Zhang, Wei Shi, Jian Xie, Shuang Li, Ruihan Yang, Tinghui Zhu, and 1 others. 2024. From persona to personalization: A survey on role-playing language agents. <i>arXiv preprint arXiv:2404.18231</i> .	706
651		707
652		708
653		709
654		710
655	Olga Chernikova, Nicole Heitzmann, Matthias Stadler, Doris Holzberger, Tina Seidel, and Frank Fischer. 2020. Simulation-based learning in higher education: A meta-analysis. <i>Review of educational research</i> , 90(4):499–541.	711
656		712
657		713
658		714
659		715
660	Yu Ying Chiu, Liwei Jiang, Bill Yuchen Lin, Chan Young Park, Shuyue Stella Li, Sahithya Ravi, Mehar Bhatia, Maria Antoniak, Yulia Tsvetkov, Vered Shwartz, and 1 others. 2024. Culturalbench: a robust, diverse and challenging benchmark on measuring (the lack of) cultural knowledge of llms.	716
661		717
662		718
663		719
664		720
665		721
	Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, and 1 others. 2021. Training verifiers to solve math word problems. <i>arXiv preprint arXiv:2110.14168</i> .	722
		723
		724
		725
		726
	Simon Cooper, Robyn Cant, Jo Porter, Fiona Bogossian, Lisa McKenna, Susannah Brady, and Stephanie Fox-Young. 2012. Simulation based learning in midwifery education: a systematic review. <i>Women and Birth</i> , 25(2):64–78.	727
		728
		729
		730
		731
		732
		733
		734
		735
		736
		737
		738
		739
		740
		741
		742
		743
		744
		745
		746
		747
		748
		749
		750
		751
		752
		753
		754
		755
		756
		757
		758
		759
		760
		761
		762
		763
		764
		765
		766
		767
		768
		769
		770
		771
		772
		773
		774
		775
		776
		777
		778
		779
		780
		781
		782
		783
		784
		785
		786
		787
		788
		789
		790
		791
		792
		793
		794
		795
		796
		797
		798
		799
		800
		801
		802
		803
		804
		805
		806
		807
		808
		809
		810
		811
		812
		813
		814
		815
		816
		817
		818
		819
		820
		821
		822
		823
		824
		825
		826
		827
		828
		829
		830
		831
		832
		833
		834
		835
		836
		837
		838
		839
		840
		841
		842
		843
		844
		845
		846
		847
		848
		849
		850
		851
		852
		853
		854
		855
		856
		857
		858
		859
		860
		861
		862
		863
		864
		865
		866
		867
		868
		869
		870
		871
		872
		873
		874
		875
		876
		877
		878
		879
		880
		881
		882
		883
		884
		885
		886
		887
		888
		889
		890
		891
		892
		893
		894
		895
		896
		897
		898
		899
		900
		901
		902
		903
		904
		905
		906
		907
		908
		909
		910
		911
		912
		913
		914
		915
		916
		917
		918
		919
		920
		921
		922
		923
		924
		925
		926
		927
		928
		929
		930
		931
		932
		933
		934
		935
		936
		937
		938
		939
		940
		941
		942
		943
		944
		945
		946
		947
		948
		949
		950
		951
		952
		953
		954
		955
		956
		957
		958
		959
		960
		961
		962
		963
		964
		965
		966
		967
		968
		969
		970
		971
		972
		973
		974
		975
		976
		977
		978
		979
		980
		981
		982
		983
		984
		985
		986
		987
		988
		989
		990
		991
		992
		993
		994
		995
		996
		997
		998
		999
		1000

720	Shreya Havaldar, Young Min Cho, Sunny Rai, and Lyle Ungar. 2025. Culturally-aware conversations: A framework & benchmark for llms. In <i>Proceedings of the Fourth Workshop on Bridging Human-Computer Interaction and Natural Language Processing (HCI+NLP)</i> , pages 220–229.	773
721		774
722		775
723		776
724		777
725		
726	Shehenaz Hossain and Haithem Afli. 2025. Craft: An explanation-based framework for evaluating cultural reasoning in multilingual language models. <i>arXiv preprint arXiv:2510.14014</i> .	778
727		779
728		780
729		781
730	Ronald Inglehart, Miguel Basanez, Jaime Diez-Medrano, Loek Halman, and Ruud Luijkx. 2000. World values surveys and european values surveys, 1981-1984, 1990-1993, and 1995-1997. <i>Ann Arbor-Michigan, Institute for Social Research, ICPSR version</i> .	782
731		783
732		784
733		785
734		786
735		787
736	Guan Jian, Junfei Wu, Jian Guan, Li Jia-nan, Jun Wu, Cheng Chuanqi, Jia-Nan Li, Wu Wei, Chuanqi Cheng, and Wei Wu. 2025. A survey on personalized alignment - the missing piece for large language models in real-world applications. https://doi.org/10.48550/arxiv.2503.17003 , pages 5313–5333.	788
737		789
738		790
739		791
740		792
741		793
742		794
743	Dongsheng Jiang, Yuchen Liu, Songlin Liu, Jin’e Zhao, Hao Zhang, Zhen Gao, Xiaopeng Zhang, Jin Li, and Hongkai Xiong. 2023. From clip to dino: Visual encoders shout in multi-modal large language models. <i>arXiv preprint arXiv:2310.08825</i> .	795
744		796
745		797
746		798
747		
748	Mohsinul Kabir, Ajwad Abrar, and Sophia Ananiadou. 2025. Break the checkbox: Challenging closed-style evaluations of cultural alignment in llms. <i>arXiv preprint arXiv:2502.08045</i> .	799
749		800
750		801
751		802
752	Dayeon Ki, Rachel Rudinger, Tianyi Zhou, and Marine Carpuat. 2025. Multiple llm agents debate for equitable cultural alignment. <i>arXiv preprint arXiv:2505.24671</i> .	803
753		804
754		805
755		
756	Min-Sun Kim. 2017. Culture-based conversational constraints theory. <i>The international encyclopedia of intercultural communication</i> , pages 1–10.	806
757		807
758		808
759	Kathrin Knutzen, Liane Rothenberger, Irina Tribusean, and Yi Xu. 2025. Using social virtual reality in teaching intercultural communication. <i>Technology, Knowledge and Learning</i> , pages 1–21.	809
760		810
761		811
762		812
763	David A Kolb. 1984. Experiential learning: Experience as the source of learning and development. <i>EC.: Prentice Hall</i> .	813
764		814
765		815
766	Louis Kwok, Michal Bravansky, and Lewis D Griffin. 2024. Evaluating cultural adaptability of a large language model via simulation of synthetic personas. <i>arXiv preprint arXiv:2408.06929</i> .	816
767		817
768		818
769		819
770	Fatimah Lateef. 2010. Simulation-based learning: Just like the real thing. <i>Journal of emergencies, trauma, and shock</i> , 3(4):348–352.	820
771		821
772		822
	Cheng Li, Mengzhuo Chen, Jindong Wang, Sunayana Sitaram, and Xing Xie. 2024a. Culturellm: Incorporating cultural differences into large language models. <i>Advances in Neural Information Processing Systems</i> , 37:84799–84838.	823
		824
	Cheng Li, Damien Teney, Linyi Yang, Qingsong Wen, Xing Xie, and Jindong Wang. 2024b. Culturepark: Boosting cross-cultural understanding in large language models. <i>Advances in Neural Information Processing Systems</i> , 37:65183–65216.	825
		826
	Chen Cecilia Liu, Iryna Gurevych, and Anna Korhonen. 2025a. Culturally aware and adapted nlp: A taxonomy and a survey of the state of the art. <i>Transactions of the Association for Computational Linguistics</i> , 13:652–689.	
	Chen Cecilia Liu, Anna Korhonen, and Iryna Gurevych. 2025b. Cultural learning-based culture adaptation of language models. <i>arXiv preprint arXiv:2504.02953</i> .	
	Chunhua Liu, Kabir Manandhar Shrestha, and Sukai Huang. 2025c. Align: Word association learning for cross-cultural generalization in large language models. <i>arXiv preprint arXiv:2508.13426</i> .	
	Reem I Masoud, Martin Ferianc, Philip Treleaven, and Miguel Rodrigues. 2025. Cultural alignment in large language models using soft prompt tuning. <i>arXiv preprint arXiv:2503.16094</i> .	
	Junho Myung, Nayeon Lee, Yi Zhou, Jiho Jin, Rifki Putri, Dimosthenis Antypas, Hsuvas Borkakoty, Eunsu Kim, Carla Perez-Almendros, Abinew Ali Ayele, and 1 others. 2024. Blend: A benchmark for llms on everyday knowledge in diverse cultures and languages. <i>Advances in Neural Information Processing Systems</i> , 37:78104–78146.	
	Omer Nacar, Serry Taiseer Sibae, Samar Ahmed, Safa Ben Atitallah, Adel Ammar, Yasser Alhabashi, Abdulrahman S Al-Batati, Arwa Alsehibani, Nour Qandos, Omar Elshehy, and 1 others. 2025. Towards inclusive arabic llms: A culturally aligned benchmark in arabic large language model evaluation. In <i>Proceedings of the First Workshop on Language Models for Low-Resource Languages</i> , pages 387–401.	
	Ahmed Njifenjou, Virgile Socal, Bassam Jabaian, and Fabrice Lefèvre. 2024. Role-play zero-shot prompting with large language models for open-domain human-machine conversation. <i>arXiv preprint arXiv:2406.18460</i> .	
	Daphna Oyserman. 2015. Culture as situated cognition. <i>Emerging trends in the social and behavioral sciences</i> , 318:1–20.	
	Ya Pang. 2022. Evaluation and promotion of a multidimensional information intelligent speech system in dialect teaching. <i>Journal of Sensors</i> .	
	Siddhesh Pawar, Junyeong Park, Jiho Jin, Arnav Arora, Junho Myung, Srishti Yadav, Faiz Ghifari	

Xuhui Zhou, Hao Zhu, Leena Mathur, Ruohong Zhang, Haofei Yu, Zhengyang Qi, Louis-Philippe Morency, Yonatan Bisk, Daniel Fried, Graham Neubig, and 1 others. 2023b. Sotopia: Interactive evaluation for social intelligence in language agents. *arXiv preprint arXiv:2310.11667*.

A Experiments and Instructions

A.1 params

Model inference was performed using the vLLM library with the following settings: temperature was set to 0.0 to disable stochasticity and ensure deterministic generation, a single output sequence was generated ($n = 1$), the maximum number of generated tokens was set to 1024, and generation was terminated upon encountering the stop token ($ID = 2$).

A.2 Training details

In this section, we provide training details of our baselines. We provide the details of the corresponding training datasets and the training parameters. All models are tested 5 times and all results we report are the average performances.

SFT_{kn} We construct the training dataset based on the extracted cultural knowledge. The dataset consists of 8551 data items, and is composed of 7 subsets, including extracted cultural knowledge QA-pairs, cultural questions from CultureScope, and general-purpose data. The quantities and proportions of these data are reported in Table 3. All data are formatted into a triplet of $(Instruction, Input, Output)$. For each knowledge instance, we convert it into a Question-Answer pair by applying Qwen2.5-72B-Instruct (Yang et al., 2024) to generate a corresponding question whose answer is the knowledge instance.

RL_{kn} We leverage the cultural knowledge QA-pairs mentioned above to perform reinforcement learning. We convert the triplet of $(Instruction, Input, Output)$ into a tuple $(scenario, knowledge)$, where *scenario* is the generated question, whose answer is the corresponding *knowledge*. In this setting, the model is instructed with:

Question: *scenario*

B Method

To investigate the effects of scenarios, roles, and perspectives on the improvement of cultural com-

Dataset	Data Quantity	Proportion
Knowledge-QA	2205	25.79%
CultureScope	800	9.36%
Alpaca	1500	17.54%
CommonsenseQA	1000	11.69%
GSM8K	800	9.36%
IFEval	846	9.89%
MMLU	1400	16.37%
Total	8551	100%

Table 3: Data counts and proportions of the constituent subsets in the SFT dataset. There are three types of data comprising the dataset: Extracted cultural knowledge QA-pairs; cultural questions from CultureScope; data for general-purpose tasks.

petence, we design several forms of interaction. As is shown in Figure 6, specifically, we construct distinct instructions that prompt the model to engage in simulated learning under different perspectives and task settings. The instructions for these interaction modes are presented as follows. In the settings, \mathcal{C} denotes the scenario context, $\mathcal{R} = r_1, r_2, \dots, r_n$ represents the set of roles, and $\mathcal{G} = g_1, g_2, \dots, g_n$ denotes the corresponding set of social goals, and \mathcal{K} denotes the cultural knowledge.

CuSiR_{mt} Under this setting, the model is instructed to role-play a character within the scenario and take actions in a turn-by-turn manner, conditioned on the scenario description, its role identity, and the accumulated dialogue history. The corresponding instruction is formatted as:

Scenario: \mathcal{C}

Cultural Background: \mathcal{K}

Social Goal: g_i

You are playing the role: r_i

Begin the conversation. Use natural dialogue and act in-character.

CuSiR_{sc} This setting simulates the role-play interaction in a planning-based or imagined manner, prompting the model to observe the scenario, roles, and objectives from an omniscient perspective and generate a complete set of possible multi-turn **short conversations** in a single pass, rather than reacting turn by turn. The corresponding instruction is formatted as:

Scenario: \mathcal{C}

Cultural Background: \mathcal{K}

Social Goal: \mathcal{G}

Generate a conversation based on the scenario and social goal. Use natural dialogue appropriate

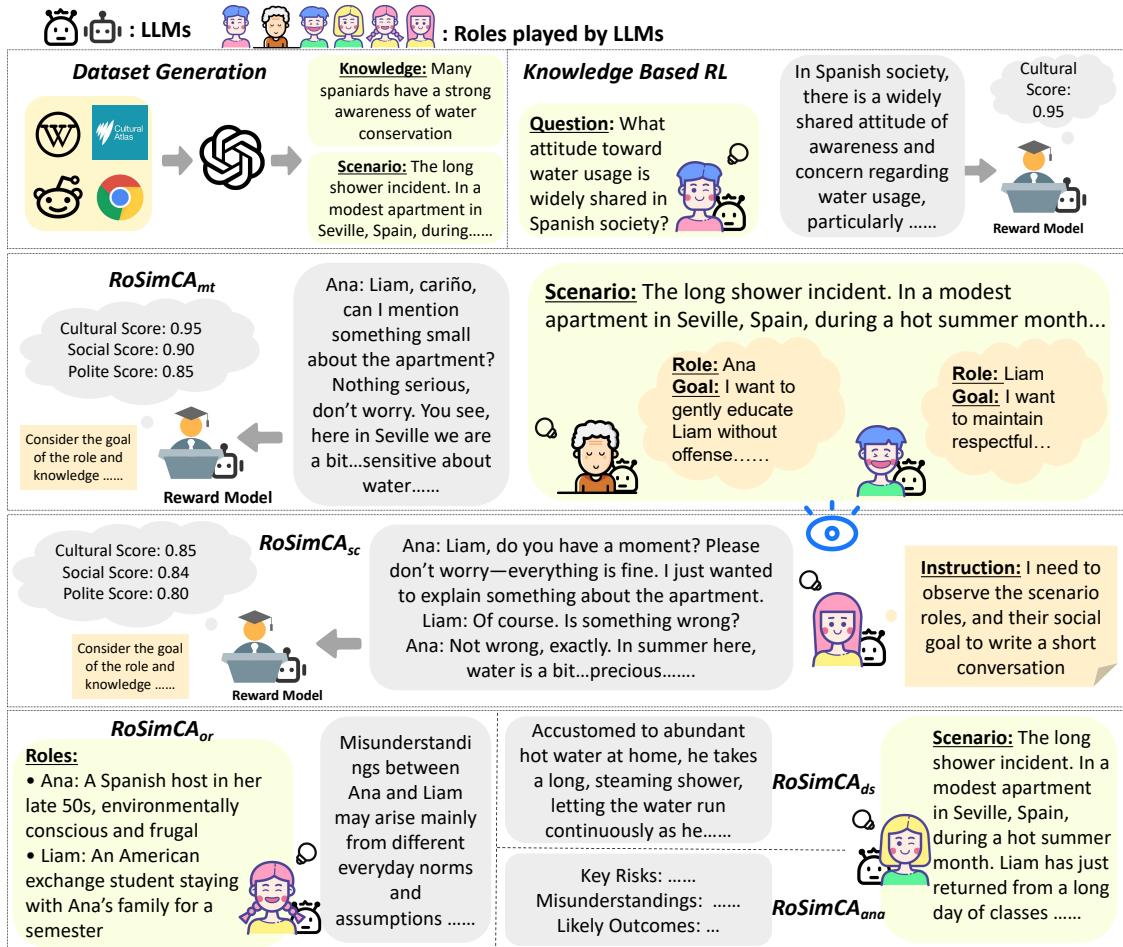


Figure 6: In different settings, distinct instructions are employed to prompt the model to engage in simulated learning under different perspectives and task settings.

to the context.

CuSiR_{ds} The model observes only the scenario and is asked to write a narrative **description** of what happens in the situation. Rather than performing interaction with specific roles, the model focuses on describing events sequentially, providing a coherent account of interactions and developments within the scenario. This method emphasizes understanding the scenario as a whole rather than generating dialogue from specific roles. The corresponding instruction is formatted as:

Scenario: \mathcal{C}

Write a narrative description of what happens in this situation.

CuSiR_{ana} In this setting, the model is asked to describe the key risks, cultural misunderstandings in the scenario, and likely outcomes. Here, the model engages in cultural reasoning and risk assessment, identifying potential pitfalls and misunder-

standings that could arise. This method highlights the model’s ability to **analyze** social and cultural dynamics from an observer’s perspective, without simulating dialogue. The corresponding instruction is formatted as:

Scenario: \mathcal{C}

Describe the key risks, cultural misunderstandings, and likely outcomes in this situation.

CuSiR_{or} The scenario is removed and **only the roles** are provided. By focusing solely on roles, this method examines the model’s ability to infer interactions and cultural dynamics based on role characteristics alone, without contextual information from the scenario, which isolates the effect of role information on cultural reasoning. The corresponding instruction is formatted as:

Roles: \mathcal{R}

Briefly describe what kinds of misunderstandings might arise between these roles in general.

C Evaluation Details

C.1 CultureScope

CultureScope is an evaluation framework and dataset generator for assessing cultural understanding in LLMs. It goes beyond fixed question sets by introducing a theoretically grounded, multi-dimensional schema for cultural knowledge classification. This approach enables systematic construction of culture-specific evaluation data across different languages and regions. In the benchmark it proposes, it contains 4 tasks: Factual: test the model on factual knowledge questions; Conceptual: test the models’ grasp of hidden cultural concepts; Mislead: test models’ ability to identify cultural bias in simulated scenarios; Multi-hop: test models’ ability of multi-hop cultural reasoning in simulated scenarios. The former two focus on the mastery of cultural knowledge, and the latter two focus on the application of these knowledge instances.

As we focus on the enhancement of cultural application abilities, we sample 1000 instances from the latter two tasks. All questions from the sampled test set are subjective questions. The evaluation of such questions is based on LLM as a judge, where a judge model is employed to judge whether the responses are correct. Due to the inherent characteristics of culture, many culture-related questions are often highly subjective and context-dependent, rarely admitting a single correct answer. Instead, responses from human annotators typically reflect broad tendencies or reasonable interpretations. Therefore, for such questions, we instructed the judge model in its prompt that a model’s response should be considered correct as long as it does not fundamentally contradict the reference answer and reasonably addresses the problem presented in the scenario. The specific prompt instructions are provided in Section H. Moreover, given that questions assessing a model’s cultural competence often involve multiple key points, we employed the DeepSeek-R1 model as our judge model, as its advanced reasoning capabilities make it better suited for handling complex, multi-faceted scenarios.

C.2 CulturalBench

CulturalBench is a benchmark by human-AI cultural teaming, designed to evaluate LLMs’ cultural competence across diverse sociocultural contexts. It focuses on annotating diverse and challenging questions for cultural knowledge. The dataset tests

how well LLMs understand culturally specific information across regions. It probes knowledge related to daily practices, norms, values, and social conventions that vary between cultures, measuring a model’s performance against human baselines and identifying weaknesses in cultural competence. It contains 17 culturally diverse topics such as food preferences, greeting etiquette, and other everyday cultural phenomena. Questions are verified by multiple independent annotators to ensure quality and explanatory depth. The benchmark is evaluated in at least two settings—multiple-choice (CulturalBench-Easy) and harder true/false variants.

We sample all questions about Spanish culture from CulturalBench-hard, and all data items are organized into a binary T/F format. For each option of a question, we ask the model to judge whether the option correctly answers the question.

C.3 BLEnD

BLEnD is a hand-crafted benchmark designed to evaluate LLMs’ everyday knowledge across diverse cultures and languages. The content of BLEnD includes quotidian topics such as local foods, traditional celebrations, sports activities in schools, work-life practices, family customs, and education. The dataset features two evaluation formats: short-answer questions, where models must generate a response to an open prompt, and multiple-choice questions, where models select the correct answer from a set of options. Each sample is grounded in culturally relevant scenarios that are familiar to native speakers but not always well documented online.

We sample all questions about Spanish culture from BLEnD and perform the following format reformulation. Each BLEnD instance consists of a question, multiple candidate answers, and the number of annotators who endorsed each answer. We transformed each instance into a multiple-choice question, where the number of annotators supporting an answer was converted into a score assigned to that option, and the sum of all option scores constituted the total score for the question. For each question, the model was required to identify all reasonable answers among the options; the score for that question was then computed as the sum of the scores corresponding to the options selected by the model. The reported accuracy is the ratio of the total scores obtained by the model across all questions to the sum of the total possible scores for

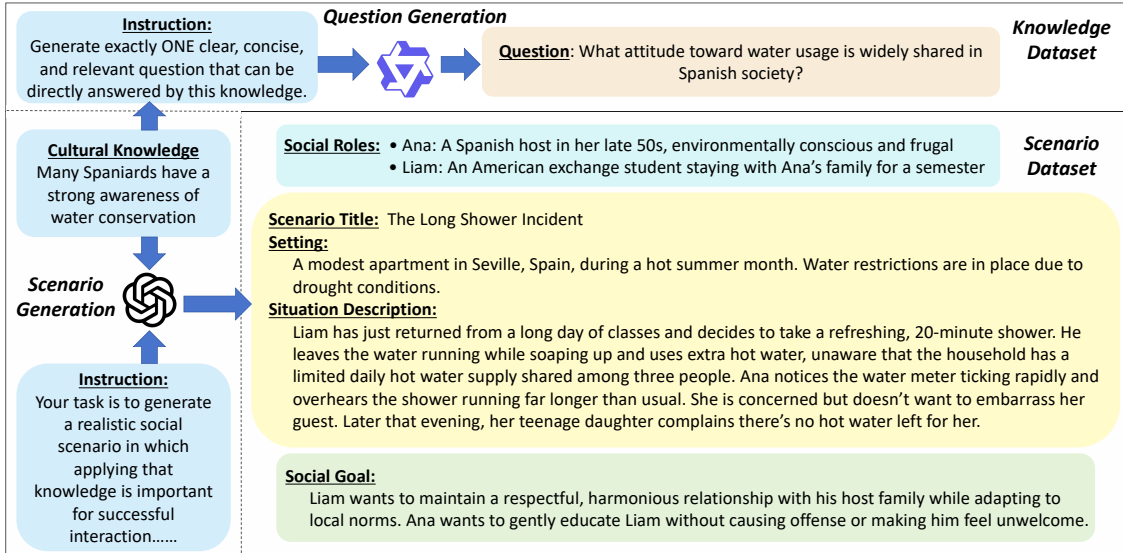


Figure 7: Examples of the generated knowledge dataset and scenario dataset. Under the guidance of extracted cultural knowledge, we prompt GPT-4o to generate simulated social scenarios in which applying the corresponding knowledge is important for successful interaction. A complete social scenario instance consists of a scenario context, roles, and each role's social goals. The scenario context is made up of a title, its setting and a paragraph of description. We further generate a question for each knowledge item to further facilitate training.

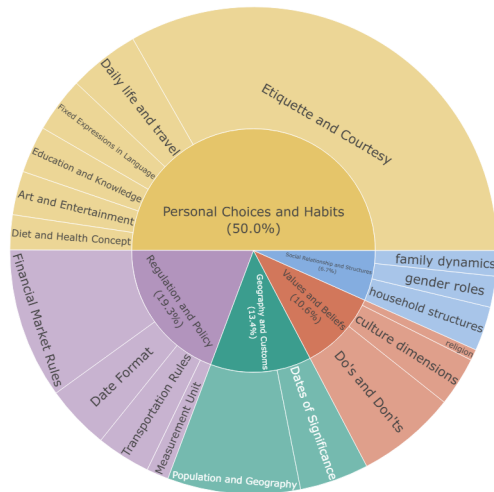


Figure 8: Dimensional distribution of extracted cultural knowledge and simulated scenario dataset.

id	Url
1	https://culturalatlas.sbs.com.au
2	https://businessculture.org
3	https://en.wikipedia.org
4	https://www.quora.com
5	https://www.expatica.com
6	https://cibtvisas.com
7	https://www.spanishpod101.com
8	https://kummuni.com
9	https://www.donquijote.org
10	https://www.enforex.com

Table 4: Data sources of the extracted cultural knowledge. These websites are mainly professional cultural websites, educational institutions and social media platforms, possessing high data quality of cultural knowledge.

all questions.

D Data Examples and Statistics

D.1 Cultural Knowledge Dataset

The extraction of cultural knowledge is Internet-based. We leverage the taxonomy of SAGE, which encompasses 100+ cultural dimensions spanning symbolic, behavioral, and value levels. We constructed retrieval queries based on these dimensional keywords, retained the top3 retrieval results from Google, and used GPT-4o to summarize and

rewrite them. We collect **2205** pieces of cultural knowledge in total. The examples are illustrated in Figure 7. We conducted a distributional analysis of the data, and the proportions of content across different dimensions are illustrated in the Figure 8. The information source of these data instances are provided in Table 4.

1171
1172
1173
1174
1175
1176
1177

1178
1179
1180
1181
1182

1183
1184

1185
1186
1187
1188

1189

1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206

1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226

D.2 Simulated Scenario Dataset

The simulated scenario dataset is generated through knowledge guidance. The distribution of this dataset is aligned with that of the knowledge dataset. The example is illustrated in Figure 7.

E Case Study: Good Situated Cultural Reasoning

In order to further explore what constitutes good situated reasoning, we conduct several empirical studies on the outputs of the Instruct model and models trained with CuSiR.

E.1 Case 1: Situated Bias Detection

Question You are planning a business trip to Spain and want to make a good impression on your Spanish colleagues. A friend who has visited Spain before gives you the following advice:

1. Spanish people love it when foreigners show off their knowledge of Spanish culture, so make sure to tell them all about the history of bullfighting and flamenco dancing.
2. Meetings in Spain are very formal, so stick strictly to the agenda and avoid any casual conversation.
3. To build trust, share personal stories and show genuine interest in their culture, but be careful not to overdo it.

Which piece of advice is most likely to lead to a cultural misunderstanding or misinterpretation? Explain why.

Response-Instruct 2. Meetings in Spain are very formal, so stick strictly to the agenda and avoid any casual conversation.

Explanation: Meeting in Spain is very formal, so stick strictly to the agenda and avoid any casual conversation. This advice is most likely to lead to a cultural misunderstanding or misinterpretation because it suggests that casual conversation is not appropriate, which may be seen as disrespectful or rude by Spanish colleagues. In Spanish culture, casual conversation is often seen as a sign of friendliness and good will, and **adhering strictly to the agenda may be seen as a lack of respect for the Spanish culture and its social norms**. Therefore, it is important to be mindful of cultural differences and to adapt one’s behavior accordingly.

1. Spanish people love it when foreigners show off their knowledge of Spanish culture, so make sure to tell them all about the history of bullfighting and flamenco dancing. This advice is likely to lead

to a good impression because it shows that you are knowledgeable about Spanish culture and are willing to share your knowledge with your colleagues. However, it may not be appropriate to tell them all about the history of bullfighting and flamenco dancing, as this may be considered offensive or inappropriate by some Spanish colleagues. Therefore, it is important to be mindful of cultural differences and to adapt one’s behavior accordingly.

3. To build trust, share personal stories and show genuine interest in their culture, but be careful not to overdo it. This advice is likely to lead to a good impression because it shows that you are interested in building trust with your colleagues and are willing to share your personal stories and show genuine interest in their culture. However, it may not be appropriate to share too many personal stories or to show too much interest in their culture, as this may be seen as a lack of respect for their privacy or as a sign of disrespect for their culture. Therefore, it is important to be mindful of cultural differences and to adapt one’s behavior accordingly.

Response-CuSiR The advice that is most likely to lead to a cultural misunderstanding or misinterpretation is: 1. Spanish people love it when foreigners show off their knowledge of Spanish culture.

This advice is problematic because it assumes that Spanish people are inherently interested in learning about foreign cultures. In reality, Spanish people are generally very proud of their own culture and history, and they may feel uncomfortable or even offended if they feel that their culture is being trivialized or reduced to mere knowledge. This could lead to a misunderstanding of the speaker’s intentions and potentially cause offense.

The other pieces of advice are more likely to be well-received and understood by Spanish colleagues:

2. Meetings in Spain are very formal, so stick strictly to the agenda and avoid any casual conversation.

This advice is appropriate because Spanish culture places a high value on formality and respect for authority. It is generally considered rude to interrupt a meeting or to engage in casual conversation during a formal meeting.

3. To build trust, share personal stories and show genuine interest in their culture, but be careful not to overdo it.

This advice is also appropriate because it ac-

1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248

1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277

1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290

1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319

1320
1321
1322
1323
1324
1325
1326
1327

knowledges the importance of building trust and showing genuine interest in others, while also respecting the cultural norms of Spain. It is a balanced approach that avoids cultural misunderstandings by being mindful of the cultural context.

In conclusion, the advice that is most likely to lead to a cultural misunderstanding or misinterpretation is the first piece of advice, as it assumes that Spanish people are inherently interested in learning about foreign cultures, which is not necessarily the case. The other pieces of advice are more likely to be well-received and understood by Spanish colleagues.

Analysis The comparison between the two responses further highlights differences in analytical quality in situated cultural reasoning. Response of the Instruct model identifies Advice 2 as potentially problematic, and this concern is not entirely unfounded, as an overly rigid avoidance of informal interaction may indeed hinder rapport-building in Spanish professional settings. However, the response reasonably notes that **an absolute avoidance of casual conversation may conflict with Spanish norms** that value interpersonal warmth and informal rapport-building. However, this insight is weakened by exaggerated claims and imprecise reasoning, particularly the assertion that adhering strictly to an agenda would be perceived as disrespectful. In practice, Spanish professional meetings often combine formal structures with limited small talk, rather than rejecting formality altogether.

By contrast, response of CuSiR offers a more nuanced and contextually grounded analysis by identifying the subtle social risk of overtly “showing off” cultural knowledge, a behavior more likely to be perceived as superficial or patronizing. This comparison suggests that higher-quality situated cultural reasoning is not merely about detecting a plausible cultural tension, but about accurately calibrating its social significance and supporting it with precise, context-aware justification.

E.2 Case 2: Applicational Cultural Reasoning

Question You are a Spanish student who has just moved to the United States for a year of study. Your American friends invite you to their Christmas party, and they mention that they will be doing a “Secret Santa” gift exchange. You are excited but also a bit confused because you are familiar with the “amigo secreto” tradition in Spain.

- 1. Describe the “amigo secreto” tradition in Spain. How does it differ from the “Secret Santa” tradition in the United States? 1328
1329
1330
- 2. Your American friends suggest that you bring a gift that is typical of your home country. What kind of gift would you choose, and why? 1331
1332
1333
- 3. How might the cultural differences in gift-giving practices between Spain and the United States affect the way you participate in the “Secret Santa” exchange? 1334
1335
1336
1337

Response-Instruct 1. The “amigo secreto” tradition in Spain involves selecting a secret friend to exchange gifts with, much like the “Secret Santa” tradition in the United States. However, the key difference is that in Spain, the gift-giving is typically done in person, often at a family gathering or a local event, whereas in the United States, **it is usually done online** or through a digital platform. 1338
1339
1340
1341
1342
1343
1344
1345

2. As a Spanish student, you would choose a gift that is typical of your home country. For example, you might choose a traditional Spanish dish, such as paella or a bottle of Spanish wine, to represent your culture. This gift would be a meaningful and authentic representation of your heritage. 1346
1347
1348
1349
1350
1351

3. The cultural differences in gift-giving practices between Spain and the United States would affect the way you participate in the “Secret Santa” exchange. In Spain, the gift-giving is a social event, and the gift itself is a reflection of your cultural identity. In the United States, gift-giving is often more individualistic, and the gift itself is a reflection of your personal preferences. Therefore, you might need to be mindful of the cultural context and adjust your gift-giving accordingly to ensure that it is both meaningful and appropriate for the exchange. 1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363

Response-CuSiR 1. The “amigo secreto” tradition in Spain is a gift-giving game where each person brings a gift for another person in the group, but the recipient does not know who they are until the exchange is complete. This tradition is a form of gift-giving that is more personal and intimate, as it involves a level of trust and friendship. In contrast, the “Secret Santa” tradition in the United States is a more impersonal gift-giving game where each person brings a gift for another person in the group, but the recipient does not know who they are until the exchange is complete. This tradition is more about competition and fun, and it is less personal. 1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377

1378 2. As a Spanish student, I would choose a gift
1379 that is typical of my home country, which is Spain.
1380 A gift that would be typical of Spain could be a
1381 traditional Spanish dish, such as paella or tapas.
1382 I would choose this gift because it is a tangible
1383 representation of my culture and would allow me
1384 to share my heritage with my American friends.

1385 3. The cultural differences in gift-giving prac-
1386 tices between Spain and the United States could
1387 affect the way I participate in the “Secret Santa” ex-
1388 change. In Spain, gift-giving is a more personal and
1389 intimate tradition, and the recipient does not know
1390 who they are until the exchange is complete. In con-
1391 trast, in the United States, gift-giving is more im-
1392 personal and competitive, and the recipient knows
1393 who they are before the exchange. As a Spanish
1394 student, I would need to be mindful of these cul-
1395 tural differences and adjust my gift-giving practices
1396 accordingly. For example, I would need to be more
1397 careful about the level of trust and friendship that I
1398 build with my American friends, and I would need
1399 to be more aware of the level of competition and
1400 fun that is involved in the exchange.

1401 This answer provides a detailed explanation of
1402 the “amigo secreto” tradition in Spain, the differ-
1403 ences between it and the “Secret Santa” tradition in
1404 the United States, and how cultural differences in
1405 gift-giving practices between Spain and the United
1406 States could affect the way a Spanish student par-
1407 ticipates in the “Secret Santa” exchange. It also
1408 provides specific examples of gifts that would be
1409 typical of Spain and how a Spanish student might
1410 need to adjust their gift-giving practices to accom-
1411 modate the cultural differences between Spain and
1412 the United States.

1413 **Analysis** A comparison of the two responses re-
1414 veals clear differences in reasoning quality and
1415 cultural understanding. Response of the Instruct
1416 model demonstrates a basic awareness of cross-
1417 cultural differences by mentioning contrasts be-
1418 tween Spain and the United States and propos-
1419 ing culturally meaningful gifts. However, its
1420 overall quality is undermined by factual inaccura-
1421 cies—most notably the incorrect characterization
1422 of Secret Santa as **primarily online**—excessive
1423 repetition, and vague overgeneralizations such as
1424 labeling U.S. gift-giving as “more individualistic”.
1425 These issues indicate shallow and imprecise cul-
1426 tural reasoning that weakens both credibility and in-
1427 terpretability. In contrast, response of CusiR model
1428 exhibits higher-quality situated cultural reasoning

1429 by providing a coherent and contextually grounded
1430 account of amigo secreto, emphasizing interper-
1431 sonal intimacy, trust, and social expectations, and
1432 contrasting these elements with the more casual
1433 and fun-oriented nature of Secret Santa in the U.S.
1434 It further connects cultural differences to concrete
1435 behavioral adaptations in gift-giving practices. De-
1436 spite minor overstatements regarding competitive-
1437 ness, response of CusiR remains factually accurate,
1438 logically structured, and analytically richer. Over-
1439 all, the comparison suggests that higher-quality
1440 responses are characterized by accuracy, contex-
1441 tual nuance, and a clear linkage between cultural
1442 understanding and practical action, whereas lower-
1443 quality responses suffer from imprecision, redun-
1444 dancy, and superficial generalization.

1445 F Rewards

1446 In this section, we provide the detailed statistics on
1447 the distributions of the training process of different
1448 reward models. Figure 9 presents the statistical
1449 results. We provide statistics for two types of re-
1450 wards: one computed as the raw sample mean, and
1451 the other smoothed using a simple moving aver-
1452 age, where each value is replaced by the mean of
1453 itself and its neighboring points within a fixed-size
1454 window. The window size is set to 3. It can be ob-
1455 served that, whether using the smoothed or the raw
1456 reward distribution, Llama3 as the reward model
1457 produces a more stable reward distribution com-
1458 pared to Mistral. This increased stability shows a
1459 positive correlation with the superior performance
1460 advantage observed after training.

1461 G Annotation of LLM Judges

1462 We conducted human annotation of the LLM Judge
1463 outputs to ensure that its judgments align with ex-
1464 pert preferences. Three annotators were asked to as-
1465 sess whether the LLM Judge correctly reflected the
1466 model’s task performance, specifically whether the
1467 model had performed reasoning accurately within
1468 the given cultural context. Information about the
1469 annotators and the annotation results are presented
1470 in Table 5. We sample 500 pieces of Judge re-
1471 sults in this process. The simple agreement of the
1472 annotators is 96.5%. All annotators are recruited
1473 within our research institute and have been appro-
1474 priately compensated for their participation. Their
1475 instruction is:

1476 You are asked to evaluate the reasonableness of
1477 the Judge’s evaluation for tasks assessing situated
1478 cultural reasoning. You will be provided with:

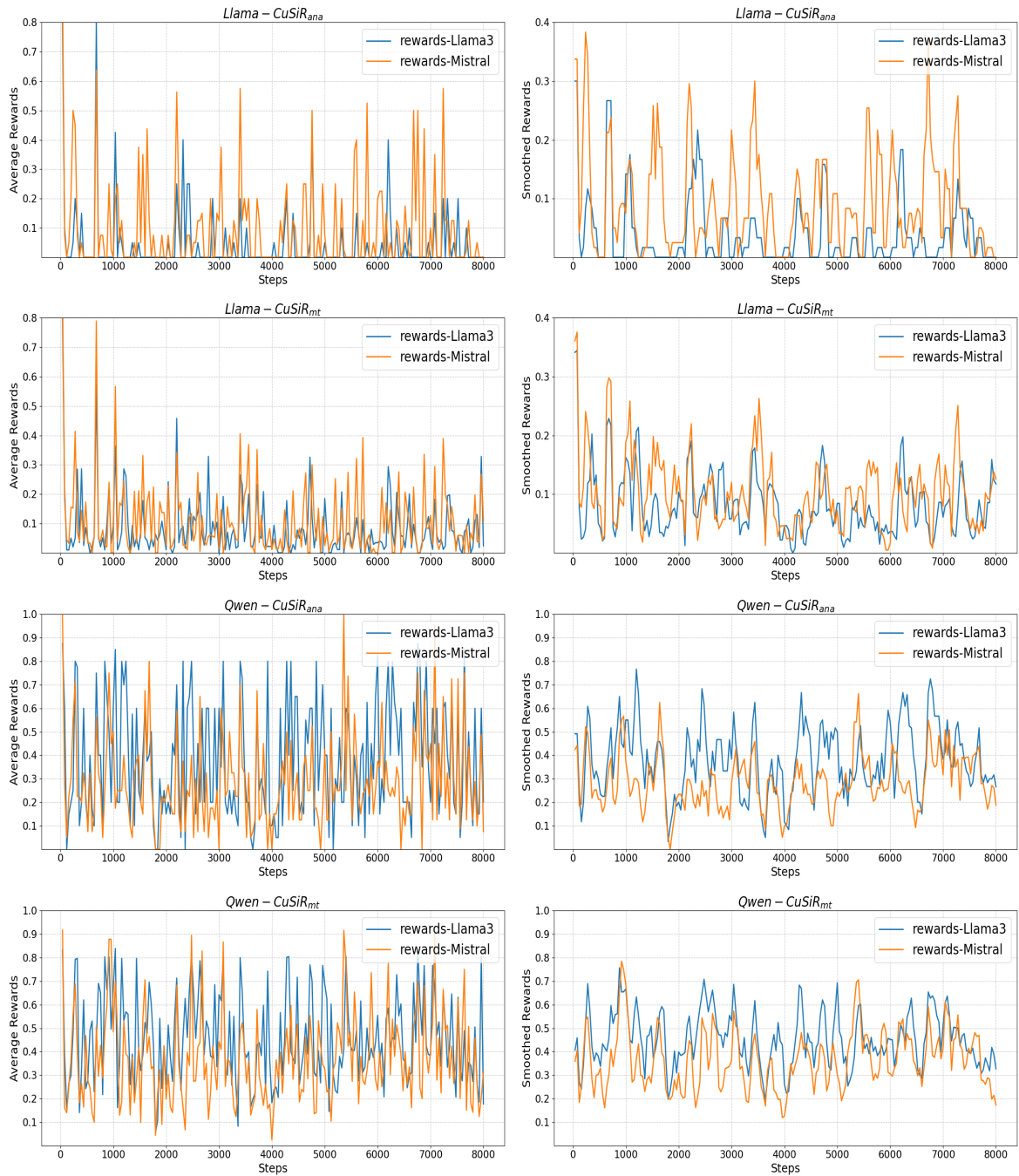


Figure 9: Distribution of the rewards generated by the two reward models. It can be observed that the reward signals provided by LLaMA-3 are more stable, exhibiting smaller fluctuations over the training steps. Even though its upper bound is not as high as that of Mistral, which suggests that more stable reward signals are positively correlated with steady improvements in cultural application capability.

1479 1. The task prompt, describing a culturally situ-
 1480 ated scenario;
 1481 2. A reference answer, representing an appropri-
 1482 ate response grounded in the cultural context;
 1483 3. The model-generated response;
 1484 4. Results provided by the LLM Judge, which
 1485 includes: an analytical comparison between the
 1486 model response and the reference answer, and a
 1487 final judgment indicating Success or Failure.
 1488 You should assess whether the LLM Judge’s anal-

ysis and final decision are reasonable and well-
 justified, given the task context and the reference
 answer. Specifically, you should consider:
 1. Whether its analysis accurately reflects the
 strengths or shortcomings of the model response
 with respect to culturally appropriate reasoning
 and action; 2. Whether the final Success / Failure
 judgment logically follows from its analysis.
 Based on these criteria, assign one of the follow-
 ing labels:

id	Demographics				Annotation
	Gender	Age	Education	Cultural Background	
1	Female	30-39	Master	Spanish	98.0%
2	Male	-29	Bachelor	Spanish & Chinese	97.4%
3	Female	-29	Master	Spanish	95%

Table 5: Annotator demographics and their annotation results. The annotators are asked to judge whether the outputs of the Judge Model are consistent with the scoring criteria.

1499 *Reasonable*: The LLM Judge’s analysis and final
1500 judgment are appropriate and justified.

1501 *Unreasonable*: The LLM Judge’s analysis is
1502 flawed, incomplete, or misaligned with the cul-
1503 tural context, or the final judgment does not follow
1504 from the analysis.

1505 **H Prompts for LLMs**

1506 During the experiments, we employed both online
1507 and offline LLMs to perform data collection, gen-
1508 eration, and related procedures. Here, we provide
1509 the prompts used at each step in Table 6.

Usage	Prompt
Knowledge Extraction	<p>I will provide a web article. Your task is to extract and summarize all key characteristics and content related to the cultural dimensions of the target culture discussed in the article.</p> <p>Instructions: Present each cultural feature under a distinct heading with the following format:</p> <p>Title: [Feature Name] Description of the feature: [Clear and concise explanation of the feature] Source of information: [Direct quote from the article; indicate the paragraph number if possible]</p> <p>Ensure that: The extracted content is accurate and faithful to the original article. The presentation is well-structured, logically organized, and easy to read. Do not create or infer information that is not present in the article.</p> <p>Article for analysis: input text</p>
Scenario Generation	<p>You are a social interaction designer specializing in intercultural communication. Given a piece of target cultural knowledge, your task is to generate a realistic social scenario in which applying that knowledge is important for successful interaction. The scenario should involve clear social goals, roles, and contextual constraints.</p> <p>Cultural knowledge: knowledge</p> <p>Please generate a scenario following this format:</p> <ul style="list-style-type: none"> - Scenario Title: - Cultural Background: - Social Roles: - Setting: - Situation Description: - Social Goal: - Key Cultural Challenges: - What could go wrong if this knowledge is not applied: - Example of an appropriate behavior or utterance: <p>Make the scenario realistic, culturally grounded, and useful for training or evaluating social reasoning models.</p>
Culture Reward	<p>You are a cultural expert.</p> <p>Given the following cultural knowledge and a response in a social dialogue, evaluate how well the response aligns with the cultural expectations and norms.</p> <p>[Cultural Background]: cultural knowledge</p> <p>[Dialogue Context]: context</p> <p>[Response]: response</p> <p>Score the cultural alignment from 0.0 (completely inappropriate or ignorant of cultural norms) to 1.0 (fully aligned with cultural expectations and sensitive to cultural context). Return only a single numeric score.</p> <p>[Score]</p>
Social Reward	<p>You are an evaluator of social interactions.</p> <p>Given a social goal and the conversation context, judge whether the model’s response helps achieve that goal.</p> <p>[Social Goal]: goal</p> <p>[Dialogue Context]: context</p> <p>[Response]: response</p> <p>Rate the degree to which the response fulfills the social goal. Score from 0.0 (fails to achieve or goes against the goal) to 1.0 (fully achieves the goal). Return only a numeric score.</p> <p>[Score]</p>

Table 6: Prompts used in the experiments.

Usage	Prompt
Polite Reward	<p>You are a politeness evaluator. Evaluate the following response for politeness, respectfulness, and appropriateness within the given social and cultural context. [Dialogue Context]: context [Response]: response Score politeness from 0.0 (impolite or offensive) to 1.0 (fully polite, respectful, and appropriate). Return only a numeric score. [Score]</p>
LLM Judge:	<p>Given the following question, reference answer, and student's answer, please determine whether the student's answer is correct compared to the reference answer. As long as the student's answer does not contain conflicts with the reference answer and reasonably covers the main ideas, it should be considered correct. If the answer is correct, return "success". If the answer is incorrect or differs from the reference answer, return failed Question: question Reference Answer: reference answer Student's Answer: test response Evaluation Result: [Provide the result: "success" or "failed"]</p>

Table 6: Prompts used in the experiments.