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

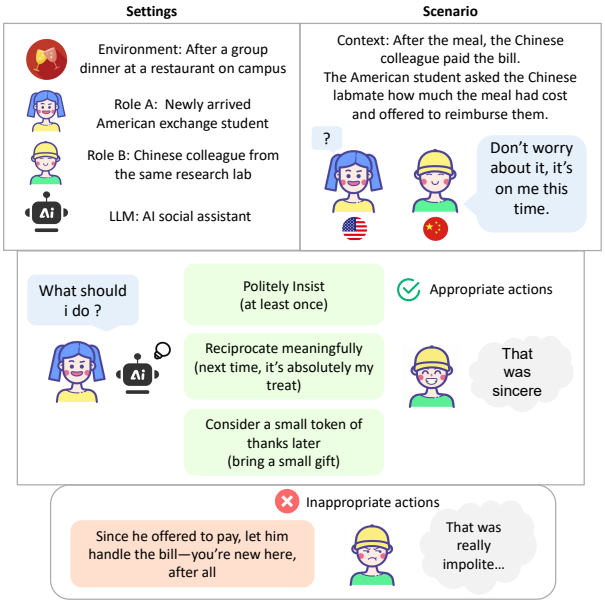

Figure 1: In real-world settings, the demand for cultural alignment is predominantly **scenario-driven**, where models are expected to appropriately **apply cultural knowledge** and norms within situated contexts.

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(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 situations (Plum et al., 2025; Hossain and Afli, 2025; Satar et al., 2025). Numerous studies have shown that improving situated cultural reason-

ing capabilities cannot rely solely on knowledge injection(Shen et al., 2024; Zhang et al., 2025; Wu et al., 2025a; Plum et al., 2025). **Situated cultural reasoning is not a knowledge acquisition problem, but a contextual application problem, and requires training signals grounded in social interactions.**

Simulation-based learning (Kolb, 1984; Lateef, 2010; Cooper et al., 2012; Aebersold, 2018; Chernikova et al., 2020) emphasizes learning through contextualized, experience-driven simulation rather than isolated factual recall. Similarly, simulation- and role-playing-based methods are widely adopted as instructional and evaluative approaches in cross-cultural research and education(Almutairi et al., 2025; Knutzen et al., 2025). Such learning paradigms are particularly well suited for tasks that require not only factual knowledge but also the ability to appropriately apply that knowledge in context; from this perspective, roles and scenarios serve as fundamental design primitives of simulation-based learning, playing a critical role in the development of cultural reasoning by enabling LLMs to practice how cultural knowledge is enacted in situated contexts.

Motivated by these studies, we propose **Cu**ltural **Si**mulation for **Si**tuated Cultural **R**easoning (CuSiR), which combines simulation and reinforcement learning(RL) to study interactions within cultural scenarios, and enables LLMs to acquire cultural knowledge and to **apply** it effectively in context. Our contributions are as follows:

1) We propose CuSiR, which is the first framework for situated cultural reasoning to integrate simulation learning with reinforcement learning by generating dialogues in synthesized scenarios to simulate real-world cultural interactions.

2) We show that combining simulation learning and RL effectively improves model performance in situated cultural reasoning scenarios.

3) Our ablation studies show that introducing simulation yields consistent performance improvements. Moreover, the observed gains stem mainly from enhanced cultural reasoning abilities rather than the introduction of new knowledge.

2 Related Work

2.1 Cultural Adaptation of LLMs

Cultural competence includes structural cultural alignment and situated cultural reasoning, and current research mainly focuses on the former. Some studies take a linguistic-level perspective(Wang et al., 2025a; Liu et al., 2025c; Dwivedi et al., 2025), focusing on how to make the outputs of LLMs conform to local etiquette and communication norms. Some works focusing on enhancing the cultural reflexivity of LLMs(Liu et al., 2025a; Li et al., 2024a,b), which enables them to critique their own Western-centric predispositions and the stereotyping or marginalization of minority and indigenous cultures. Improving axiological consistency in subjective value judgments(AlKhamissi et al., 2024; Masoud et al., 2025; Feng et al., 2025b) is another important topic. These works aim to ensure that responses to ethical, moral, and social questions reflect the normative framework of the target culture rather than a universal standard.

Close to ours, CLCA (Liu et al., 2025b) leverages simulated social interactions for supervised fine-tuning, showing effectiveness in aligning models to diverse cultural values as measured by World Values Survey(Inglehart et al., 2000) benchmarks. Their results suggest that social interaction and intent understanding during this process is important for value alignment across various cultures. Value alignment does not guarantee appropriate action selection in concrete social situations. Our work focuses on utilizing role-play strategies in simulated scenarios to enhance cultural application abilities, inspired by simulation learning theories from learning sciences and cognitive psychology.

2.2 Role-play and Simulation of LLMs

Early studies show that LLMs are effective at role-playing (Shanahan et al., 2023; Wang et al., 2024a; Njifenjou et al., 2024), and have leveraged persona simulation to evaluate and enhance their cultural capabilities(Kwok et al., 2024; AlKhamissi et al., 2024; Ki et al., 2025). However, this line of work largely remains at the level of prompt engineering and does not explicitly investigate situated cultural reasoning. LLMs are also capable of performing data synthesis based on role-play and scenarios (Li et al., 2024a; Chen et al., 2024; Wang et al., 2025b). Algorithmically generated datasets can mitigate limitations of real data and thereby 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 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: the perspectives, goals, and constraints under which the agent operates; , enabling learning from multiple viewpoints 3. interaction and feedback: how actions are evaluated, ranging from rewards to outcome-based signals.

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.

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, aiming to demonstrate that it can operate effectively across diverse cultural dimensions. Specifically, all simulated interactions and evaluations are grounded in Spanish cultural norms, which we choose for their strongly context-dependent social structures.

4.1 Scenario Data Generation

Cultural Knowledge We first extracted cultural knowledge $\mathcal{K}$ from multiple Internet-based sources. Following the taxonomy in SAGE (Guo et al., 2025), which encompasses over 100 cultural keywords spanning symbolic, behavioral, and value levels, we construct retrieval queries based on these keywords, retained the top 3 results from Google, and used GPT-4o to summarize them. The sources and the prompts used are listed in the Appendix D, with primary sources including reliable references such as Culture Atlas[1] and Wikipedia[2].

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, $\mathcal{R} = r_1, r_2, \ldots, r_n$ represents the set of roles, and $\mathcal{G} = g_1, g_2, \ldots, g_n$ 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)

We incorporate RL to leverage the scenario dataset $\mathcal{D}_s = \{\mathcal{S}_i \mid i = 1, 2, \ldots, N\}$ for training. By alignment with our preliminary framework, this approach integrates all the foundational pillars of Simulation Learning: $\mathcal{C}$ establishes environment fidelity, $\mathcal{R}$ and $\mathcal{G}$ constitute the role and state, and the RL training loop drives the interaction and feedback. Ultimately, this unified process adapts simulation-based training to effectively enhance the cultural understanding capabilities of LLMs.

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;

[1] https://culturalatlas.sbs.com.au/
[2] https://en.wikipedia.org/

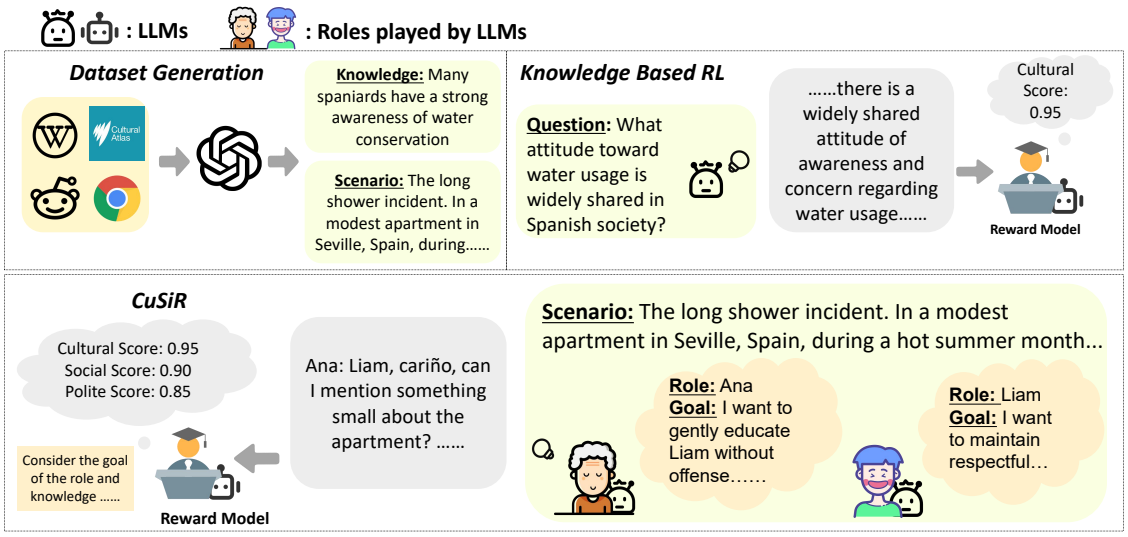

Figure 2: The overall framework of CuSiR. We leverage cultural knowledge from online educational institutions and construct simulated social scenarios for training. We design several interaction strategies, in which models are required to play roles and act as instructed, and a reward model is employed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

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.

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

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 effectiveness 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 models' cultural competence using three datasets, and details are provided in Appendix C. 1. CultureScope(Zhang et al., 2025) comprises four tasks to evaluate LLMs' cultural knowledge and cultural application. We selected two tasks focusing situated cultural reasoning: scenario-based bias identification and multi-turn cultural reasoning. 2. CulturalBench(Chiu et al., 2024) evaluates 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. We select a subset of this dataset and reformulated it as multiple-choice questions.

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 generated outputs and the reference answer. 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:

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

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 $D_{\text{KL}}(P \,\|\, Q)$ is the KL divergence:

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

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

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

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 expriments on Instruct models as they exhibit stronger instruction-following behavior and more stable task completion. All experiments were conducted on NVIDIA A800 GPUs.

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 do not control what the model learns, but when different signals become dominant. They are set as:

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

We train CuSiR with Proximal Policy Optimization (PPO), where each scenario $\mathcal{S}$ serves as the environment: the policy generates a response or action conditioned on its assigned role $\mathcal{R}$ and instruction $\mathcal{I}$, and is updated based on the composite reward defined in Sec 4.2. Implementation details and hyperparameters are provided in Appendix A.

5.3 Baselines

Knowledge-based SFT We construct the training dataset $\mathcal{D}_k = \{(q_i, a_i) \mid i = 1, 2, \ldots, N\}$ based on the extracted cultural knowledge. For each knowledge instance a_i, we apply Qwen2.5-72B-Instruct (Yang et al., 2024) to generate a corresponding question q_i 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.

Interaction-based SFT In the data synthesis phase, we leverage cultural knowledge as prompts to guide GPT-4o in generating the gold-standard behavior for each instance of S, which constitutes the dataset $\mathcal{D}_s$ for interactive behaviors.

Knowledge-based RL We conduct RL on $\mathcal{D}_k$ to remove 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 $\mathcal{D}_s$, we instruct the models to perform role-play interactions under different settings to study how roles, scenarios, and role-playing modes affect situated cultural reasoning. We conduct RL experiments on the following settings. **CuSiR$_{\text{mt}}$** is a multi-turn dialogue-based setting with turn-by-turn responses; **CuSiR$_{\text{sc}}$** adopts an omniscient,

Models / Performace		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	0.702	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
	$SFT_{\mathcal{D}_k}$	0.458(+3.3%)	0.526	0.460	0.375	0.520	0.456	0.344	**0.557**
	$SFT_{\mathcal{D}_s}$	0.466(+4.1%)	0.512	0.400	0.433	0.533	0.493	0.327	0.529
	$RL_{\mathcal{D}_k}$	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	0.557
	$CuSiR_{sc}$	**0.523(+9.8%)**	0.594	0.485	0.415	0.733	0.610	0.331	0.516
	$CuSiR_{ds}$	0.518(+9.3%)	0.587	0.415	**0.469**	0.733	**0.640**	**0.356**	0.541
	$CuSiR_{ana}$	0.509(+8.4%)	**0.594**	0.386	0.467	0.733	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
	$SFT_{\mathcal{D}_k}$	0.643 (+0.9%)	0.662	0.551	0.627	0.800	0.713	0.769	0.506
	$SFT_{\mathcal{D}_s}$	0.646(+1.2%)	0.646	0.559	0.635	0.813	0.720	0.760	0.477
	$RL_{\mathcal{D}_k}$	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
	$CuSiR_{ds}$	0.689 (+5.5%)	0.744	0.585	0.663	0.867	0.765	0.769	0.541
	$CuSiR_{ana}$	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	0.591	**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 cultural 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.

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

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 ΔJS 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 improvement on cultural knowledge QA benchmarks. 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.

Gold-standard behaviors alone is insufficient for developing situated cultural reasoning. Despite being trained on gold-standard behavioral outputs from the same scenario dataset used by CuSiR, $SFT_{\mathcal{D}_s}$ achieves only marginal improvement over $SFT_{\mathcal{D}_k}$ on CultureScope (+4.1% vs. +3.3% for Llama3; +1.2% vs. +0.9% for Qwen2.5), and is comparable to or below $RL_{\mathcal{D}_k}$ (0.466 vs. 0.465 on Llama3; 0.646 vs. 0.656 on Qwen2.5). This gap is

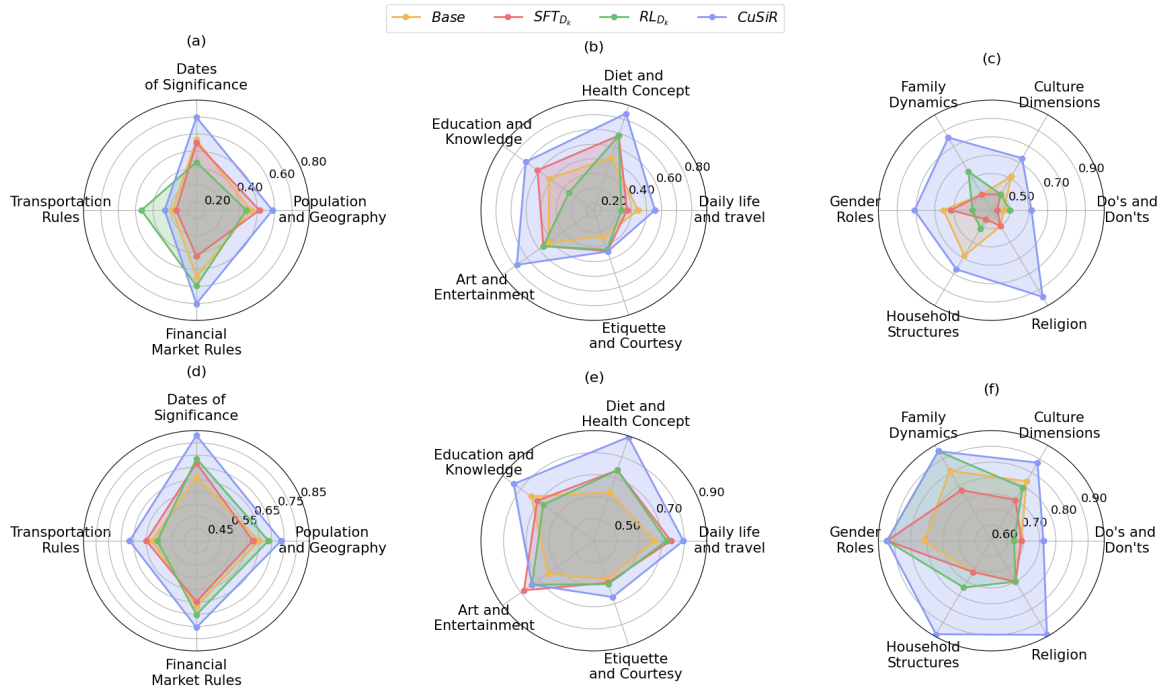

Figure 3: Results on multi cultural dimensions of Llama3(a-c) and Qwen2.5(d-f). It can be observed that, whether for explicit cultural elements such as institutional rules and individual behavioral practices, or implicit cultural aspects such as family structures and religious beliefs and taboos, CuSiR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

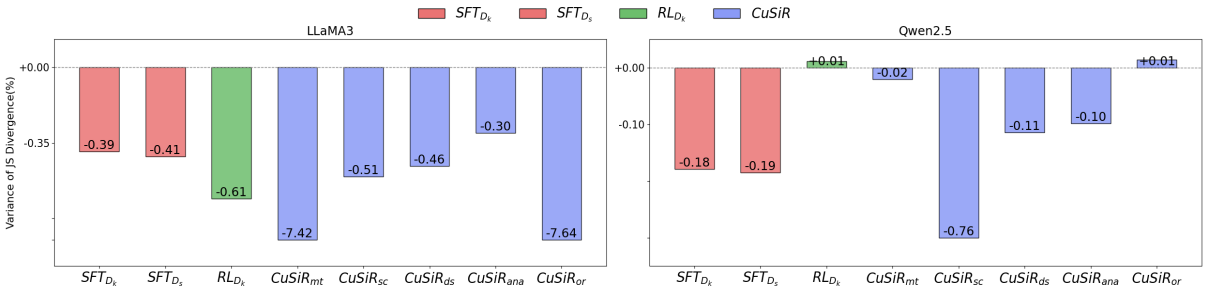

Figure 4: $\Delta\text{JS} = \text{JS}(M_{\text{trained}}) - \text{JS}(M_{\text{base}})$ against the base model, where larger values indicate closer alignment with the golden reasoning patterns. We observe that CuSiR encourages the output distribution to more closely fit reasonable situated cultural reasoning, providing further evidence of its effectiveness.

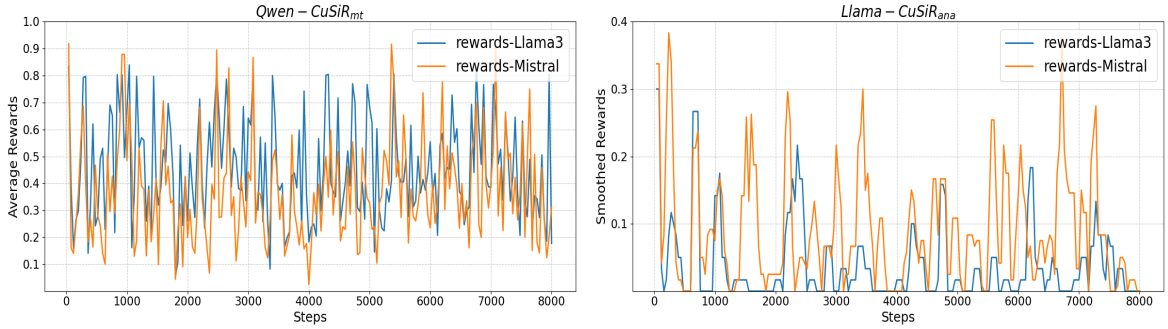

Figure 5: Distribution of the rewards generated by the two reward models. It can be observed that the reward signals provided by LLaMA-3 are more stable, exhibiting smaller fluctuations over the training steps. Even though its upper bound is not as high as that of Mistral, which suggests that more stable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

Model	Method	R_ϕ	CultureScope	Reward Statistics				
				mean	variance	std	max	min
Llama-1B	$CuSiR_{mt}$	Llama3	0.516	0.08315	0.01075	0.10369	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	0.10879	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	0.18857	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	0.19896	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 models and find that the observed gap is largely consistent with differences in their reward statistics.

substantial compared to CuSiR (+9.8% and +7.4%, respectively), suggesting that supervised imitation of gold behaviors alone is insufficient for developing situated cultural reasoning, and interaction-driven feedback through RL is essential.

6.2 Discussions

Reinforcement Learning enhances cultural reasoning. 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 RL itself can facilitate better mastery of pre-existing knowledge.

Standard supervision underperforms on implicit cultural norms. The performance gap between CuSiR and non-CuSiR baselines is largest on the CultureScope dimensions that reflect implicit cultural knowledge—social structures and underlying values. On Llama3-1B, non-CuSiR baselines fall *below* the Instruct level on both SR&S (-11.1%) and V&B (-3.4%), while CuSiR variants gain +9.0% and +8.9% respectively, yielding cross-method gaps of +20.1 and +12.3 points. On Qwen2.5-1.5B, SR&S and V&B are the only two dimensions on which non-CuSiR baselines fail to consistently improve over Instruct. These results indicate that simulation-based training is particularly effective at teaching the nuanced, implicit aspects of culture that resist standard propositional supervision, which can even erode them.

More stable reward signals lead to better performance. Table 2 reports results using different Instruct models as R_ϕ. Models trained with Llama3-rewards consistently outperform Mistral-rewards by 2–4% accuracy, particularly for the 1.5B backbone. Statistics on their reward distributions reveal that reward models with a higher effective lower bound and reduced variance across training steps yield more reliable improvements, even when their maximum reward values are not the highest. This suggests that, for cultural learning, avoiding frequent low- or zero-reward feedback is more beneficial than occasionally providing strong positive signals. Complete reward distributions are provided in Appendix F.

High-quality situated cultural reasoning(SCR) is context-sensitive, anticipatory, and practically applicable We observe that high-quality cultural reasoning is characterized by context sensitivity, anticipatory capacity, and practical applicability: rather than relying on superficial norms, it identifies subtle social cues, predicts potential misunderstandings, and guides culturally appropriate, socially effective decision-making, emphasizing relevance and actionable strategies rather than length. Detailed analysis is 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. As the first work focusing on SCR tasks, we demonstrate that both scenario data and reinforcement learning feedback are effective for cultural reasoning, providing a reference for further improvements in future research.

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, including two native Spanish speakers and one PhD student specializing in Spanish with study-abroad experience. 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 results for research purposes. The details are shown in Appendix G.

Risks To mitigate the risk of stereotyping or overgeneralization, 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 Empiri-*

cal Methods in Natural Language Processing, pages 18063–18090.

Michelle Aebersold. 2018. Simulation-based learning: No longer a novelty in undergraduate education. *Online Journal of Issues in Nursing*, 23(2).

Guillaume Alinier and Denis Oriot. 2022. Simulation-based education: deceiving learners with good intent. *Advances in Simulation*, 7(1):8.

Badr AlKhamissi, Muhammad ElNokrashy, Mai AlKhamissi, and Mona Diab. 2024. Investigating cultural alignment of large language models. *arXiv preprint arXiv:2402.13231*.

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. *BMC nursing*, 24(1):1390.

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 *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)*, pages 7580–7617.

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.

John Seely Brown, Allan Collins, and Paul Duguid. 1989. Situated cognition and the culture of learning. *1989*, 18(1):32–42.

Che-Jui Chang, Danrui Li, Seonghyeon Moon, and Mubbasir Kapadia. 2024. On the equivalency, substitutability, and flexibility of synthetic data. *arXiv preprint arXiv:2403.16244*.

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. *arXiv preprint arXiv:2404.18231*.

Olga Chernikova, Nicole Heitzmann, Matthias Stadler, Doris Holzberger, Tina Seidel, and Frank Fischer. 2020. Simulation-based learning in higher education: A meta-analysis. *Review of educational research*, 90(4):499–541.

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.

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. *arXiv preprint arXiv:2110.14168*.

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. *Women and Birth*, 25(2):64–78.

Rasha Dorgham, Lamyaa Bin Obaid, Rasha Dorgham, and Lamyaa Bin Obaid. 2025. Using ai-driven large language models (llms) in english as a foreign language education. *https://doi.org/10.21608/jfeb.2025.401861.2190*, 36(143.1):1–22.

Edgar Dubourg, Valentin Thouzeau, and Nicolas Baumard. 2024. A step-by-step method for cultural annotation by llms. *Frontiers in Artificial Intelligence*, 7:1365508.

Ashutosh Dwivedi, Siddhant Shivdutt Singh, and Ashutosh Modi. 2025. Eticor++: Towards understanding etiquettical bias in llms. *arXiv preprint arXiv:2506.08488*.

Ruixiang Feng, Shen Gao, Xiuying Chen, Lisi Chen, and Shuo Shang. 2025a. Culfit: A fine-grained cultural-aware llm training paradigm via multilingual critique data synthesis. *arXiv preprint arXiv:2505.19484*.

S Feng, WC Chan, S Chouhan, JFG Ayala, S Medicherla, K Clark, and M Shi. 2025b. Whispers of many shores: Cultural alignment through collaborative cultural expertise [publisher: arxiv].

Maxwell Forbes, Jena D Hwang, Vered Shwartz, Maarten Sap, and Yejin Choi. 2020. Social chemistry 101: Learning to reason about social and moral norms. *arXiv preprint arXiv:2011.00620*.

Usman Ghani. 2016. Role of feedbacks in simulation-based learning. *International Journal of Educational and Pedagogical Sciences*, 10(1):283 – 290.

Alastair C Graham and Sean McAleer. 2018. An overview of realist evaluation for simulation-based education. *Advances in Simulation*, 3(1):13.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Shiwei Guo, Sihang Jiang, Qianxi He, Yanghua Xiao, Jiaqing Liang, Bi Yude, Minggui He, Shimin Tao, and Li Zhang. 2025. Do large language models truly understand cross-cultural differences? *arXiv preprint arXiv:2512.07075*.

Shreya Havaldar, Young Min Cho, Sunny Rai, and Lyle Ungar. 2025. Culturally-aware conversations: A framework & benchmark for llms. In *Proceedings of the Fourth Workshop on Bridging Human-Computer Interaction and Natural Language Processing (HCI+ NLP)*, pages 220–229.

Shehenaz Hossain and Haithem Afli. 2025. Craft: An explanation-based framework for evaluating cultural reasoning in multilingual language models. *arXiv preprint arXiv:2510.14014*.

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. *Ann Arbor-Michigan, Institute for Social Research, ICPSR version*.

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.

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. *arXiv preprint arXiv:2310.08825*.

Mohsinul Kabir, Ajwad Abrar, and Sophia Ananiadou. 2025. Break the checkbox: Challenging closed-style evaluations of cultural alignment in llms. *arXiv preprint arXiv:2502.08045*.

Dayeon Ki, Rachel Rudinger, Tianyi Zhou, and Marine Carpuat. 2025. Multiple llm agents debate for equitable cultural alignment. *arXiv preprint arXiv:2505.24671*.

Min-Sun Kim. 2017. Culture-based conversational constraints theory. *The international encyclopedia of intercultural communication*, pages 1–10.

Kathrin Knutzen, Liane Rothenberger, Irina Tribusean, and Yi Xu. 2025. Using social virtual reality in teaching intercultural communication. *Technology, Knowledge and Learning*, pages 1–21.

David A Kolb. 1984. Experiential learning: Experience as the source of learning and development. *EC.: Prentice Hall*.

Louis Kwok, Michal Bravansky, and Lewis D Griffin. 2024. Evaluating cultural adaptability of a large language model via simulation of synthetic personas. *arXiv preprint arXiv:2408.06929*.

Fatimah Lateef. 2010. Simulation-based learning: Just like the real thing. *Journal of emergencies, trauma, and shock*, 3(4):348–352.

Cheng Li, Mengzhuo Chen, Jindong Wang, Sunayana Sitaram, and Xing Xie. 2024a. Culturellm: Incorporating cultural differences into large language models. *Advances in Neural Information Processing Systems*, 37:84799–84838.

Cheng Li, Damien Teney, Linyi Yang, Qingsong Wen, Xing Xie, and Jindong Wang. 2024b. Culturepark: Boosting cross-cultural understanding in large language models. *Advances in Neural Information Processing Systems*, 37:65183–65216.

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. *Transactions of the Association for Computational Linguistics*, 13:652–689.

Chen Cecilia Liu, Anna Korhonen, and Iryna Gurevych. 2025b. Cultural learning-based culture adaptation of language models. *arXiv preprint arXiv:2504.02953*.

Chunhua Liu, Kabir Manandhar Shrestha, and Sukai Huang. 2025c. Align: Word association learning for cross-cultural generalization in large language models. *arXiv preprint arXiv:2508.13426*.

Reem I Masoud, Martin Ferianc, Philip Treleaven, and Miguel Rodrigues. 2025. Cultural alignment in large language models using soft prompt tuning. *arXiv preprint arXiv:2503.16094*.

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. *Advances in Neural Information Processing Systems*, 37:78104–78146.

Omer Nacar, Serry Taiseer Sibaee, 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 *Proceedings of the First Workshop on Language Models for Low-Resource Languages*, pages 387–401.

Ahmed Njifenjou, Virgile Sucal, Bassam Jabaian, and Fabrice Lefèvre. 2024. Role-play zero-shot prompting with large language models for open-domain human-machine conversation. *arXiv preprint arXiv:2406.18460*.

Daphna Oyserman. 2015. Culture as situated cognition. *Emerging trends in the social and behavioral sciences*, 318:1–20.

Ya Pang. 2022. Evaluation and promotion of a multidimensional information intelligent speech system in dialect teaching. *Journal of Sensors*.

Siddhesh Pawar, Junyeong Park, Jiho Jin, Arnav Arora, Junho Myung, Srishti Yadav, Faiz Ghifari Haznitrama, Inhwa Song, Alice Oh, and Isabelle Augenstein. 2025. Survey of cultural awareness in language models: Text and beyond. *Computational Linguistics*, pages 1–96.

Alistair Plum, Anne-Marie Lutgen, Christoph Purschke, and Achim Rettinger. 2025. Identity-aware large language models require cultural reasoning. *arXiv preprint arXiv:2510.18510*.

Abhinav Rao, Akhila Yerukola, Vishwa Shah, Katharina Reinecke, and Maarten Sap. 2024. Normad: A framework for measuring the cultural adaptability of large language models. *arXiv preprint arXiv:2404.12464*.

Nikta Gohari Sadr, Sahar Heidariasl, Karine Megerdoomian, Laleh Seyyed-Kalantari, and Ali Emami. 2025. We politely insist: Your llm must learn the persian art of taarof. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 1819–1838.

Maarten Sap, Saadia Gabriel, Lianhui Qin, Dan Jurafsky, Noah A Smith, and Yejin Choi. 2020. Social bias frames: Reasoning about social and power implications of language. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 5477–5490.

Burak Satar, Zhixin Ma, Patrick Amadeus Irawan, Wilfried Ariel Mulyawan, Jing Jiang, Ee-Peng Lim, and Chong-Wah Ngo. 2025. Seeing culture: A benchmark for visual reasoning and grounding. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 22238–22254.

Murray Shanahan, Kyle McDonell, and Laria Reynolds. 2023. Role play with large language models. *Nature*, 623(7987):493–498.

Siqi Shen, Lajanugen Logeswaran, Moontae Lee, Honglak Lee, Soujanya Poria, and Rada Mihalcea. 2024. Understanding the capabilities and limitations of large language models for cultural commonsense. *arXiv preprint arXiv:2405.04655*.

Siqi Shen, Mehar Singh, Lajanugen Logeswaran, Moontae Lee, Honglak Lee, and Rada Mihalcea. 2025. Revisiting llm value probing strategies: Are they robust and expressive? In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 131–145.

Claudia Strauss and Naomi Quinn. 1997. *A cognitive theory of cultural meaning*, volume 9. Cambridge University Press.

Henry Tari, Nojus Sereiva, Rishabh Kaushal, Thales Bertaglia, and Adriana Iamnitchi. 2025. Towards high-fidelity synthetic multi-platform social media datasets via large language models. *arXiv preprint arXiv:2505.02858*.

Noah Wang, Zy Peng, Haoran Que, Jiaheng Liu, Wangchunshu Zhou, Yuhan Wu, Hongcheng Guo, Ruitong Gan, Zehao Ni, Jian Yang, and 1 others. 2024a. Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 14743–14777.

Qihan Wang, Shidong Pan, Tal Linzen, and Emily Black. 2025a. Multilingual prompting for improving llm generation diversity. *arXiv preprint arXiv:2505.15229*.

Ruiyi Wang, Haofei Yu, Wenxin Zhang, Zhengyang Qi, Maarten Sap, Graham Neubig, Yonatan Bisk, and Hao Zhu. 2024b. Sotopia-pi: Interactive learning of socially intelligent language agents. *arXiv preprint arXiv:2403.08715*.

Xiaoyang Wang, Hongming Zhang, Tao Ge, Wenhao Yu, Dian Yu, and Dong Yu. 2025b. Opencharacter: Training customizable role-playing llms with large-scale synthetic personas. *arXiv preprint arXiv:2501.15427*.

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi. 2023. Self-instruct: Aligning language models with self-generated instructions. In *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: long papers)*, pages 13484–13508.

Ina Wechsung. 2014. *An Evaluation Framework for Multimodal Interaction*. An Evaluation Framework for Multimodal Interaction.

Jincenzi Wu, Jianxun Lian, Dingdong Wang, and Helen Meng. 2025a. Socialcc: Interactive evaluation for cultural competence in language agents. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 33242–33271.

Meng-Chen Wu, Si-Chi Chin, Tess Wood, Ayush Goyal, and Narayanan Sadagopan. 2025b. Incorporating diverse perspectives in cultural alignment: Survey of evaluation benchmarks through a three-dimensional framework. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 17037–17072.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, and 1 others. 2024. Qwen2. 5 technical report. *arXiv preprint arXiv:2412.15115*.

Jinghao Zhang, Sihang Jiang, Shiwei Guo, Shisong Chen, Yanghua Xiao, Hongwei Feng, Jiaqing Liang, Minggui HE, Shimin Tao, and Hongxia Ma. 2025. Culturescope: A dimensional lens for probing cultural understanding in llms. *arXiv preprint arXiv:2509.16188*.

Jeffrey Zhou, Tianjian Lu, Swaroop Mishra, Siddhartha Brahma, Sujoy Basu, Yi Luan, Denny Zhou,

and Le Hou. 2023a. Instruction-following evaluation for large language models. *arXiv preprint arXiv:2311.07911*.

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$_{\mathbf{kn}}$ 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$_{\mathbf{kn}}$ 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: $scenario$

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.

B Method

To investigate the effects of scenarios, roles, and perspectives on the improvement of cultural competence, 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, $\mathcal{R} = r_1, r_2, \ldots, r_n$ represents the set of roles, and $\mathcal{G} = g_1, g_2, \ldots, g_n$ denotes the corresponding set of social goals, and $\mathcal{K}$ denotes the cultural knowledge.

CuSiR$_{\mathbf{mt}}$ 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 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$_{\mathbf{sc}}$ 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 instruction is formatted as:

Scenario: $\mathcal{C}$

Cultural Background: $\mathcal{K}$

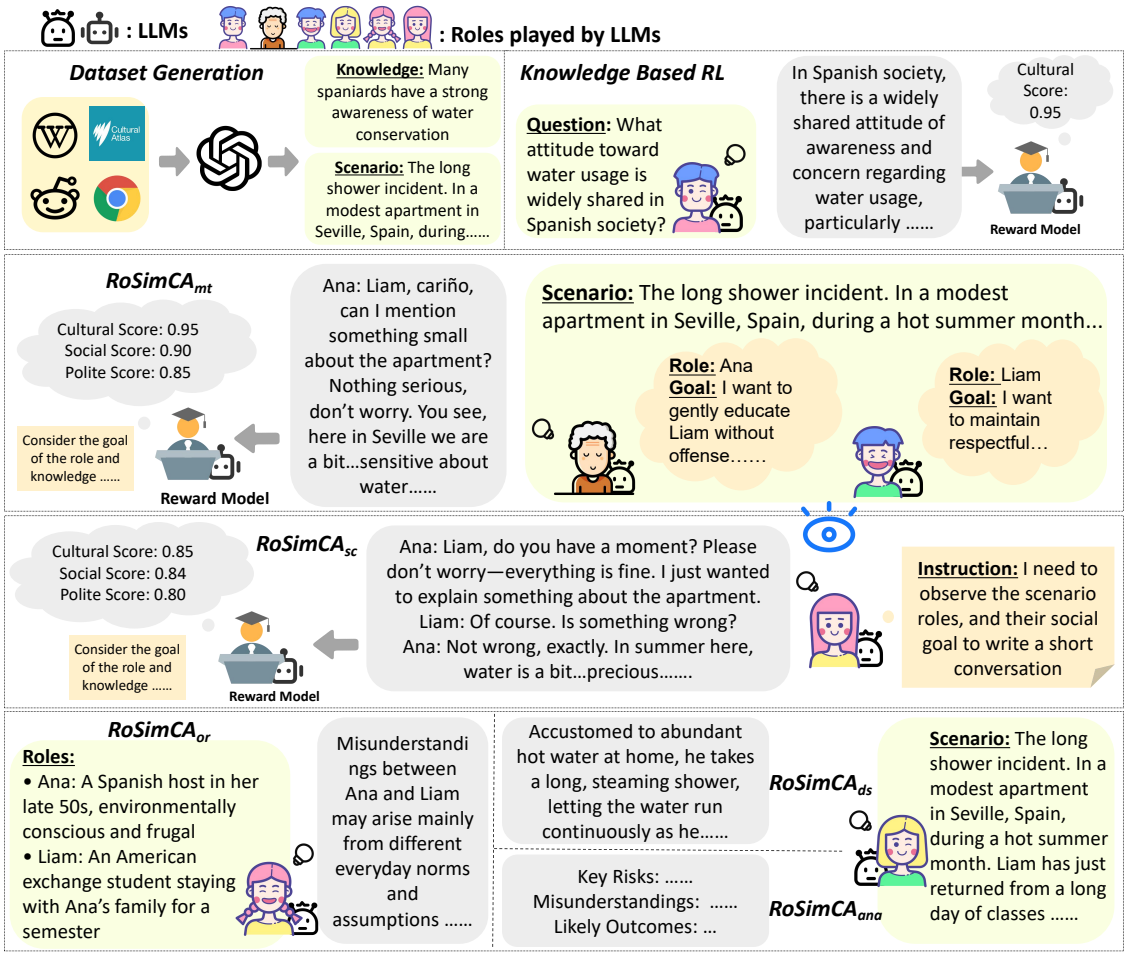

Figure 6: In different settings, distinct instructions are employed to prompt the model to engage in simulated learning under different perspectives and task settings.

Social Goal: $\mathcal{G}$

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

CuSiR$_{\mathbf{ds}}$ 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 instruction is formatted as:

Scenario: $\mathcal{C}$

Write a narrative description of what happens in this situation.

CuSiR$_{\mathbf{ana}}$ 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 misunderstandings 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 instruction is formatted as:

Scenario: $\mathcal{C}$

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

CuSiR$_{\mathbf{or}}$ 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 instruction is formatted as:

Roles: $\mathcal{R}$

Briefly describe what kinds of misunderstand-

ings 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

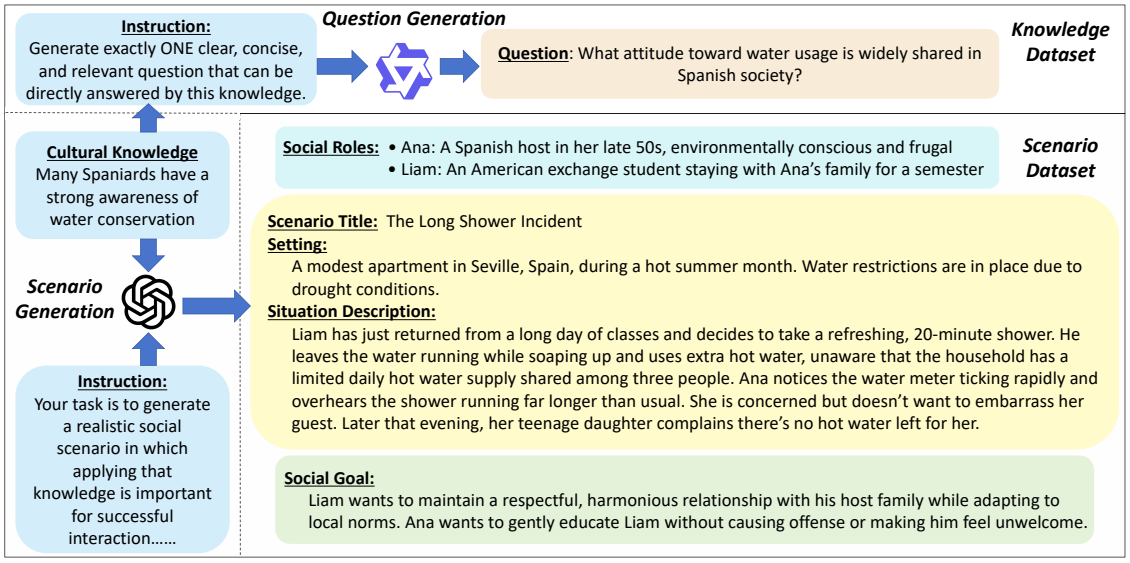

Figure 7: Examples of the generated knowledge dataset and scenario dataset. Under the guidance of extracted cultural knowledge, we prompt GPT-4o to generate simulated social scenarios in which applying the corresponding knowledge is important for successful interaction. A complete social scenario instance 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.

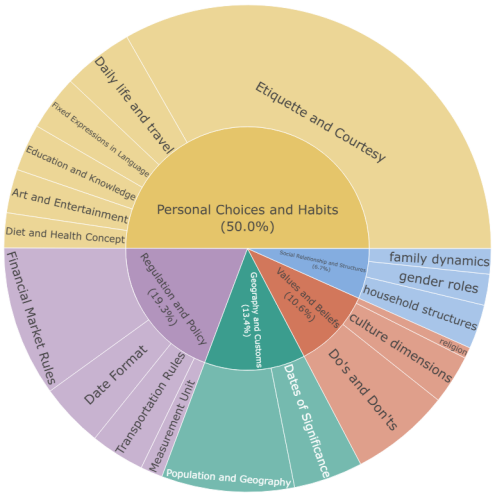

Figure 8: Dimensional distribution of extracted cultural knowledge and simulated scenario dataset.

the total scores obtained by the model across all questions to the sum of the total possible scores for 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 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.

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

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.

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 acknowledges 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?

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?

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?

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.

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.

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.

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.

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

3. The cultural differences in gift-giving practices between Spain and the United States could affect the way I participate in the "Secret Santa" exchange. In Spain, gift-giving is a more personal and intimate tradition, and the recipient does not know who they are until the exchange is complete. In contrast, in the United States, gift-giving is more impersonal and competitive, and the recipient knows who they are before the exchange. As a Spanish student, I would need to be mindful of these cultural differences and adjust my gift-giving practices accordingly. For example, I would need to be more careful about the level of trust and friendship that I build with my American friends, and I would need to be more aware of the level of competition and fun that is involved in the exchange.

This answer provides a detailed explanation of the "amigo secreto" tradition in Spain, the differences between it and the "Secret Santa" tradition in the United States, and how cultural differences in gift-giving practices between Spain and the United States could affect the way a Spanish student participates in the "Secret Santa" exchange. It also provides specific examples of gifts that would be typical of Spain and how a Spanish student might need to adjust their gift-giving practices to accommodate the cultural differences between Spain and the United States.

Analysis A comparison of the two responses reveals clear differences in reasoning quality and cultural understanding. Response of the Instruct model demonstrates a basic awareness of cross-cultural differences by mentioning contrasts between Spain and the United States and proposing culturally meaningful gifts. However, its overall quality is undermined by factual inaccuracies—most notably the incorrect characterization of Secret Santa as **primarily online**—excessive repetition, and vague overgeneralizations such as labeling U.S. gift-giving as "more individualistic". These issues indicate shallow and imprecise cultural reasoning that weakens both credibility and interpretability. In contrast, response of CusiR model exhibits higher-quality situated cultural reasoning by providing a coherent and contextually grounded account of amigo secreto, emphasizing interpersonal intimacy, trust, and social expectations, and contrasting these elements with the more casual and fun-oriented nature of Secret Santa in the U.S. It further connects cultural differences to concrete behavioral adaptations in gift-giving practices. Despite minor overstatements regarding competitiveness, response of CusiR remains factually accurate, logically structured, and analytically richer. Overall, the comparison suggests that higher-quality responses are characterized by accuracy, contextual nuance, and a clear linkage between cultural understanding and practical action, whereas lower-quality responses suffer from imprecision, redundancy, and superficial generalization.

F Rewards

In this section, we provide the detailed statistics on the distributions of the training process of different reward models. Figure 9 presents the statistical results. We provide statistics for two types of rewards: one computed as the raw sample mean, and the other smoothed using a simple moving average, where each value is replaced by the mean of itself and its neighboring points within a fixed-size window. The window size is set to 3. It can be observed that, whether using the smoothed or the raw reward distribution, Llama3 as the reward model produces a more stable reward distribution compared to Mistral. This increased stability shows a positive correlation with the superior performance advantage observed after training.

G Annotation of LLM Judges

We conducted human annotation of the LLM Judge outputs to ensure that its judgments align with expert preferences. Three annotators were asked to assess whether the LLM Judge correctly reflected the model's task performance, specifically whether the model had performed reasoning accurately within the given cultural context. Information about the annotators and the annotation results are presented in Table 5. We sample 500 pieces of Judge results in this process. The simple agreement rate of the annotators is 96.5%, reflecting the proportion of items on which the annotators assigned the same labels. All annotators are recruited within our research institute and have been appropriately com-

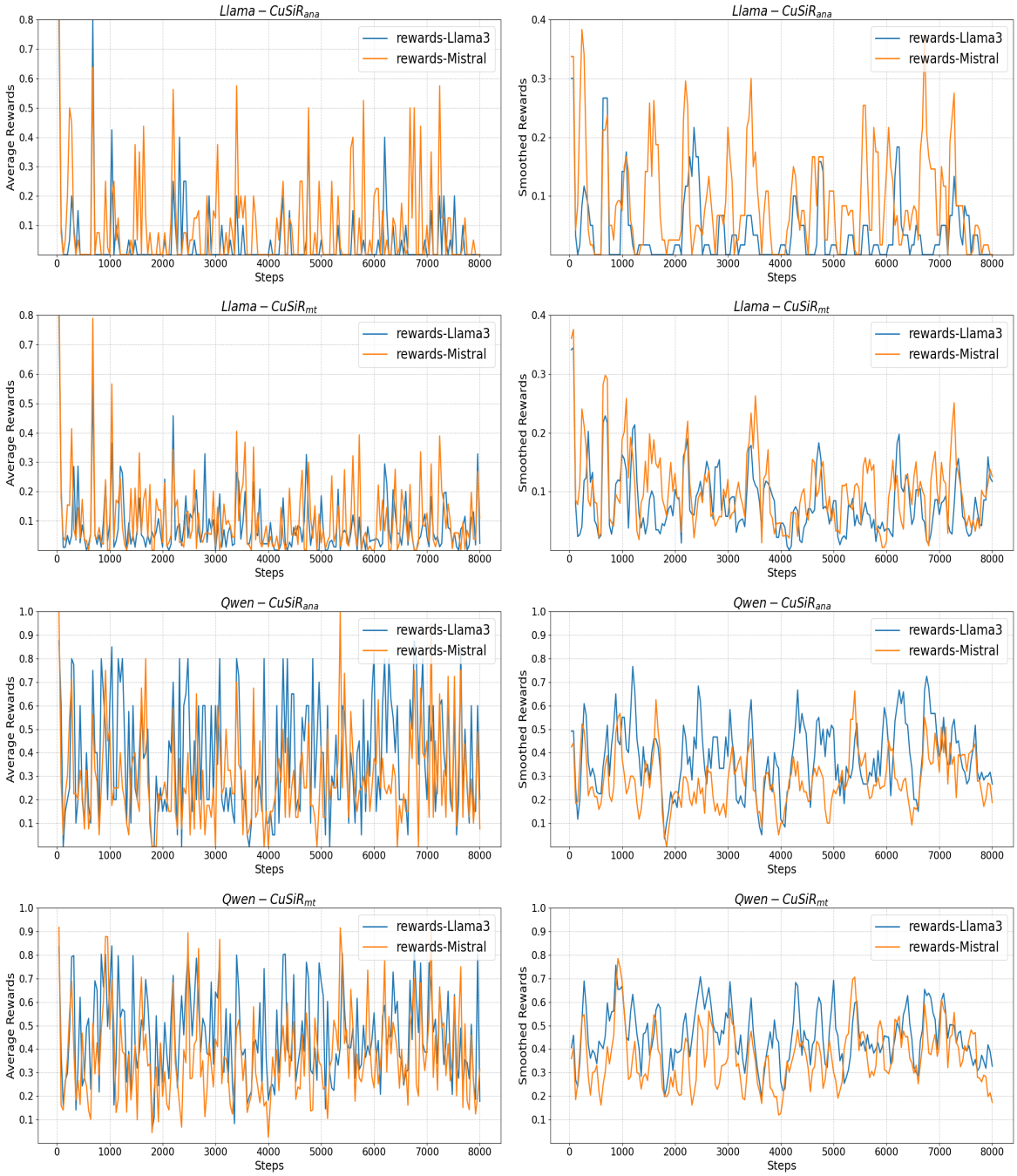

Figure 9: Distribution of the rewards generated by the two reward models. It can be observed that the reward signals provided by LLaMA-3 are more stable, exhibiting smaller fluctuations over the training steps. Even though its upper bound is not as high as that of Mistral, which suggests that more stable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

pensated for their participation. Their instruction is:

> You are asked to evaluate the reasonableness of the Judge's evaluation for tasks assessing situated cultural reasoning. You will be provided with:
>
> 1. The task prompt, describing a culturally situated scenario;
>
> 2. A reference answer, representing an appropriate response grounded in the cultural context;
>
> 3. The model-generated response;
>
> 4. Results provided by the LLM Judge, which includes: an analytical comparison between the model response and the reference answer, and a final judgment indicating Success or Failure.
>
> You should assess whether the LLM Judge's analysis 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

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.

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 following labels:

Reasonable: The LLM Judge's analysis and final judgment are appropriate and justified.

Unreasonable: The LLM Judge's analysis is flawed, incomplete, or misaligned with the cultural context, or the final judgment does not follow from the analysis.

H Prompts for LLMs

During the experiments, we employed both online and offline LLMs to perform data collection, generation, and related procedures. Here, we provide the prompts used at each step in Table 6.

Usage	Prompt
Knowledge Extraction	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. Instructions: Present each cultural feature under a distinct heading with the following format: 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] 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. Article for analysis: input text
Scenario Generation	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. Cultural knowledge: knowledge Please generate a scenario following this format: - 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: Make the scenario realistic, culturally grounded, and useful for training or evaluating social reasoning models.
Culture Reward	You are a cultural expert. Given the following cultural knowledge and a response in a social dialogue, evaluate how well the response aligns with the cultural expectations and norms. [Cultural Background]: cultural knowledge [Dialogue Context]: context [Response]: response 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. [Score]
Social Reward	You are an evaluator of social interactions. Given a social goal and the conversation context, judge whether the model's response helps achieve that goal. [Social Goal]: goal [Dialogue Context]: context [Response]: response 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. [Score]

Table 6: Prompts used in the experiments.

Usage	Prompt
Polite Reward	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]
LLM Judge:	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"]

Table 6: Prompts used in the experiments.