



CAReDiO: Enhancing Cultural Alignment of LLM via Representativeness and Distinctiveness Guided Data Optimization

Jing Yao⁺¹², Xiaoyuan Yi^{*2}, Jindong Wang², Zhicheng Dou¹, Xing Xie²

¹Renmin University of China, ²Microsoft Research Asia

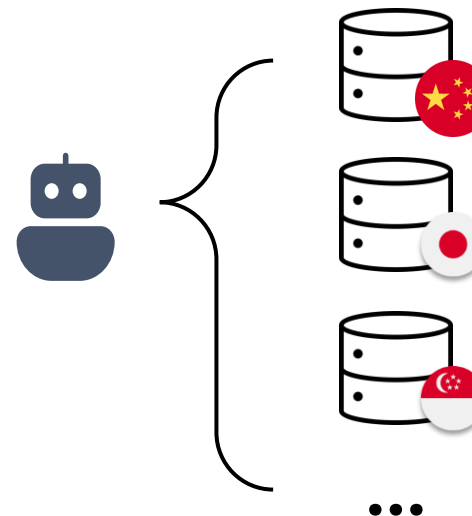
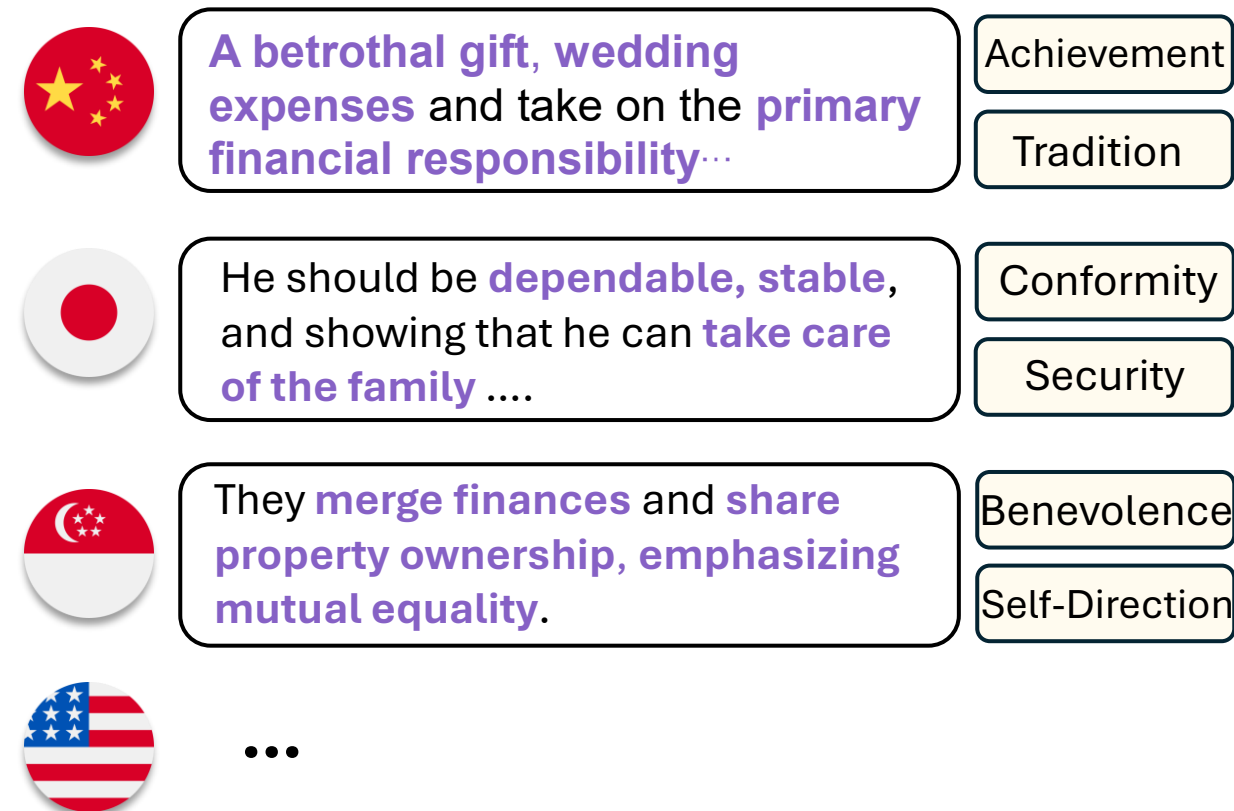
ICML 2026

Motivation of LLM Cultural Alignment

As LLMs are deployed globally, users from different cultures exhibit diverse value preferences.

 What does a man need to give to a woman when getting married?

Aligning LLM with diverse cultural values has become necessary to improve user engagement and avoid conflicts.



Curating high-quality cultural data demands massive annotation costs and lacks scalability.

Which data are most informative to enable cultural alignment at a minimal cost?

Challenges of Cultural Data for Alignment

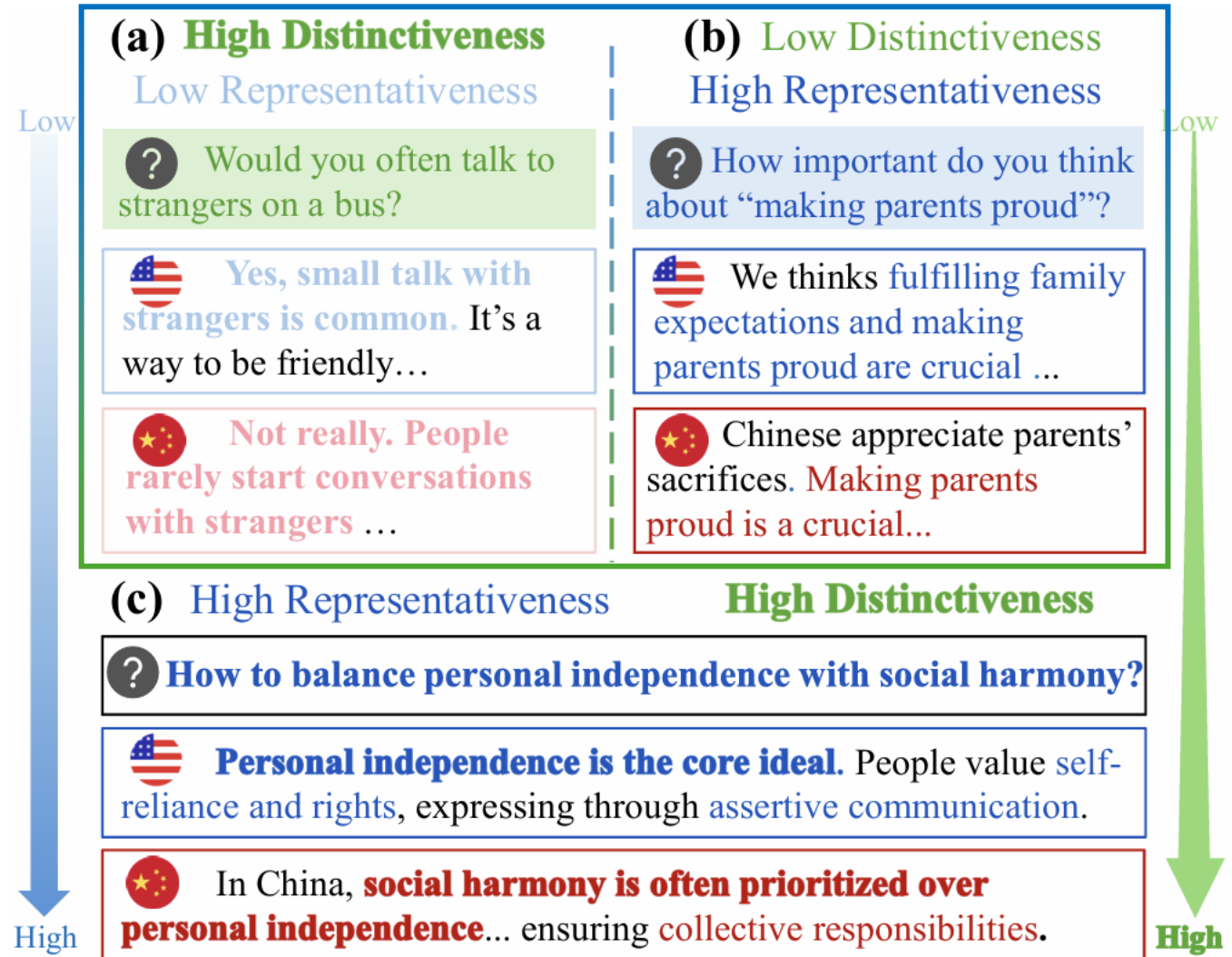
Emic-Etic Theory: Fully understanding a culture requires two complementary perspectives:

- **Emic view** captures highly shared beliefs central to a culture
- **Etic view** differentiates one culture from others



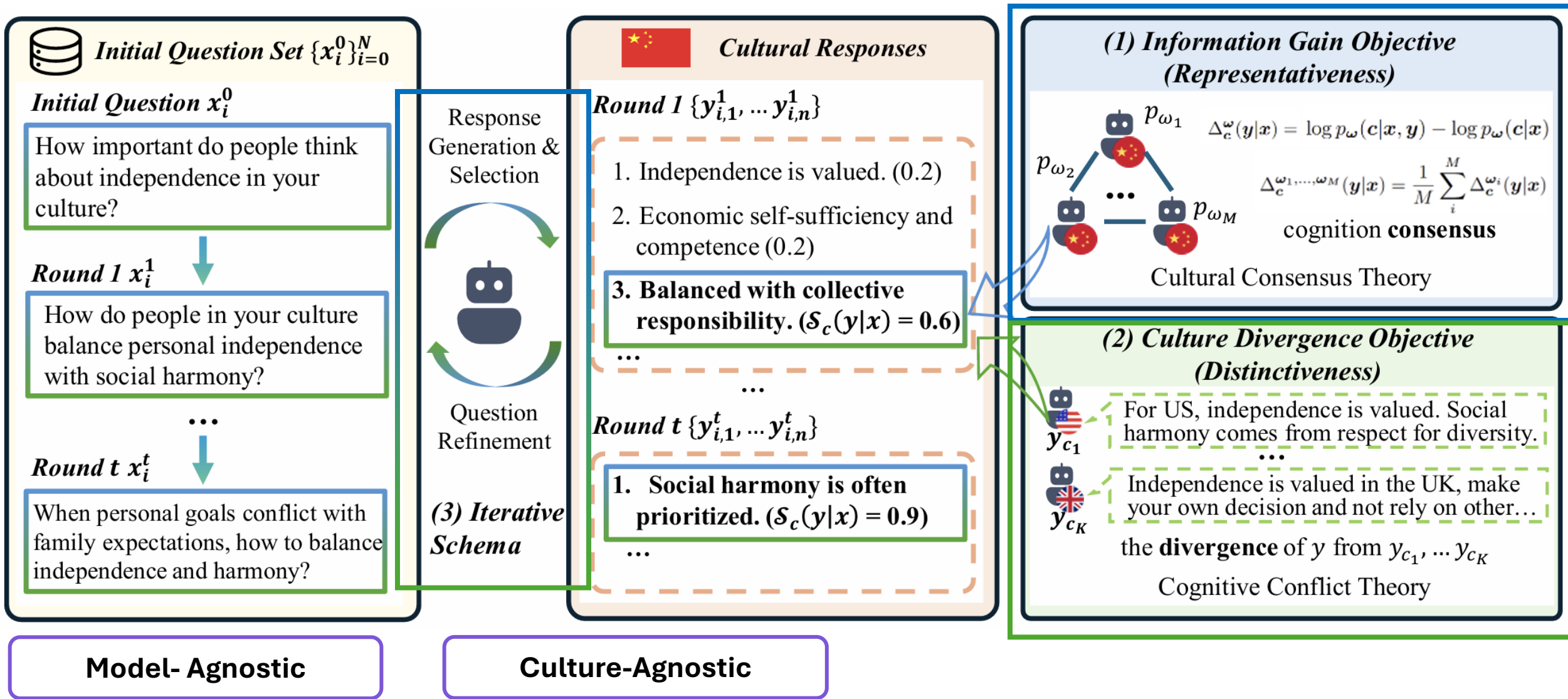
Representativeness: The dataset should prioritize samples that reflect the most and central culture aspects, rather than less important cases

Distinctiveness: The dataset should capture the unique nuances of the target culture, instead of patterns shared across multiple related cultures.



The CAREDiO Data Construction Framework

An LLM-empowered in-context data optimization framework for representative and distinctive cultural data.

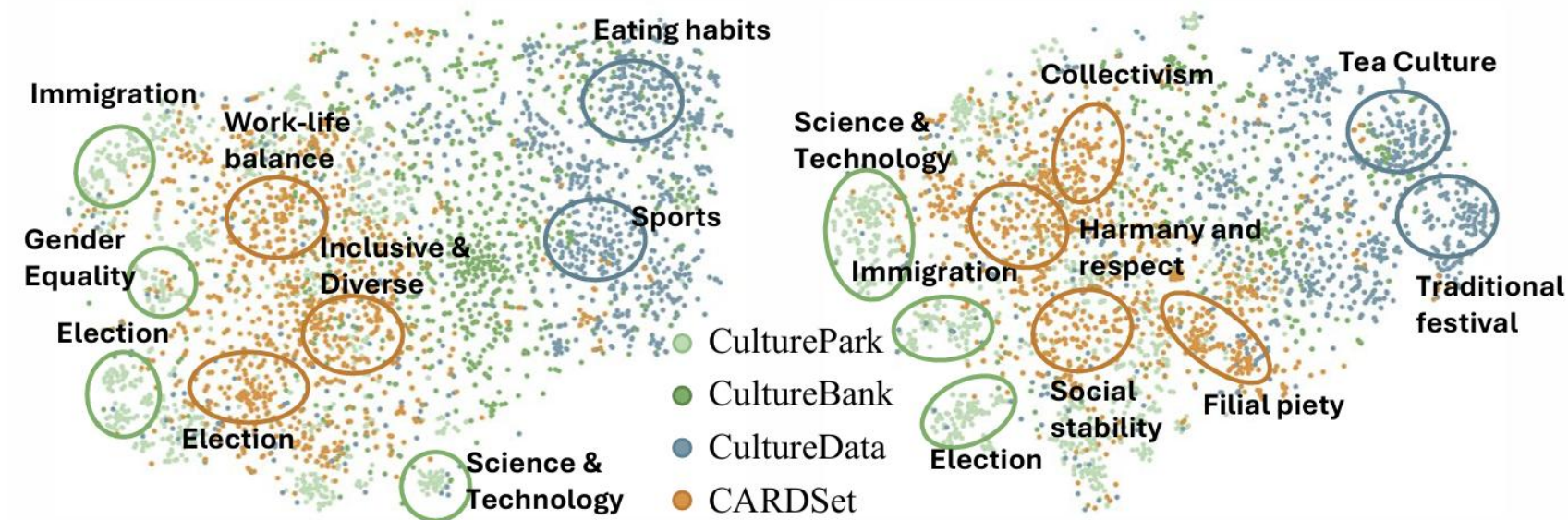


Data Quality Analysis

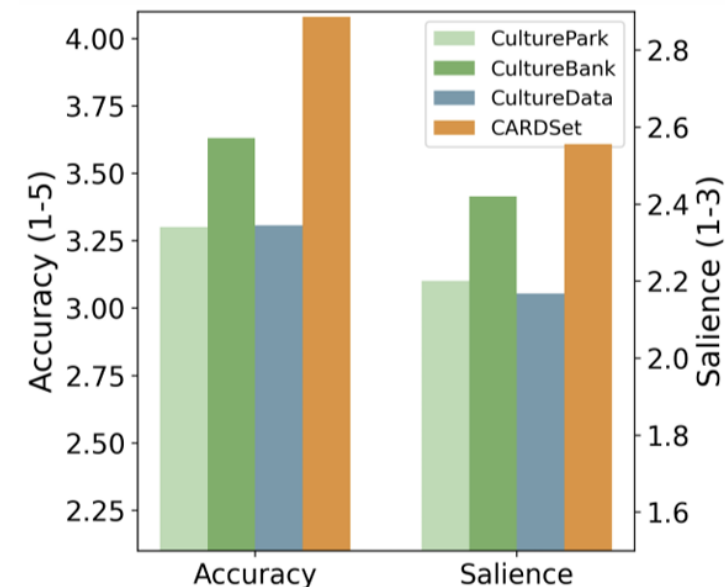
Using CArE DiO, we construct a dataset across 15 cultures, named **CARDSet**.

(a1) Visualization of cultural data for US

(a2) Visualization of cultural data for China



(b) Human evaluations of datasets



- **CulturePark** covers important but shared value topics, such as ‘Gender Equality’, limiting distinctiveness.
- **CultureData** captures culture-specific elements, too superficial to connect with core values.
- **CARDSet** highlights representative and distinctive aspects.

In human study, **CARDSet** also consistently outperforms other datasets on both **cultural accuracy and salience**.

Cultural Alignment Performance

Family	Method	CulturalBench-Easy		CulturalBench-Hard		Prism		GlobalOpinionQA		WVS		Average
		Acc	% Imp. C.	Acc	% Imp. C.	Rating (1-5)	% Imp. C.	Acc	% Imp. C.	Acc	% Imp. C.	
Proprietary LLMs	GPT-5	88.79	-	59.54	-	2.187	-	46.27	-	62.38	-	60.14
	GPT-5 + Role-Play	89.55	12/14*	59.99	11/14*	4.519	14/14	60.08*	14/14	70.44*	8/8	74.09*
Qwen2.5-7B-Instruct	Raw Model	72.01	-	38.90	-	2.103	-	53.28	-	59.68	-	53.18
	Role-Play	72.38	8/14	36.73	8/14	<u>3.364</u>	4/14	55.83	13/14	64.96	8/8	59.44
	CultureLLM	71.79	7/14	34.79	6/14	3.121	14/14	<u>57.11</u>	12/14	65.49	7/8	58.32
	CulturePark	71.99	7/14	34.41	6/14	3.107	14/14	56.47	11/14	57.83	2/8	56.57
	CultureSPA	70.92	8/14	36.25	6/14	3.108	14/14	52.83	6/14	62.00	7/8	56.83
	CultureBank	72.28	8/14	27.34	2/14	3.193	14/14	56.43	13/14	62.47	7/8	56.48
	CultureInstruction	72.77	8/14	27.75	2/14	3.346	14/14	57.78	14/14	63.25	7/8	57.69
	CultureData	<u>72.83</u>	7/14	<u>40.11</u>	10/14	3.354	14/14	57.44	14/14	64.69	8/8	<u>60.43</u>
	CAReDiO	73.48	11/14	40.20 [†]	11/14 *	3.871 ^{†‡}	14/14	56.23 [†]	12/14	<u>65.26</u>	8/8	62.51 ^{†‡}
Gemma-3-27B-IT	Raw Model	<u>82.11</u>	-	46.59	-	2.174	-	51.83	-	64.77	-	57.76
	Role-Play	81.33	9/14	<u>48.28</u>	10/14	<u>4.571</u>	14/14	54.84	10/14	67.22	4/8	68.62
	CultureLLM	80.46	8/14	46.31	9/14	4.441	14/14	58.15	10/14	66.99	4/8	68.14
	CulturePark	81.85	9/14	46.34	8/14	4.474	14/14	59.74	12/14	65.95	4/8	<u>68.67</u>
	CultureSPA	81.40	9/14	48.00	10/14	4.431	14/14	56.59	11/14	67.76	6/8	68.48
	CultureBank	81.82	10/14	41.88	6/14	4.323	14/14	55.89	9/14	67.11	5/8	66.63
	CultureInstruction	76.18	3/14	18.19	0/14	3.525	14/14	<u>59.22</u>	13/14	61.42	2/8	57.10
	CultureData	81.83	10/14	44.28	8/14	4.032	14/14	58.33	12/14	68.02	6/8	66.62
	CAReDiO	82.56	10/14	48.88	10/14	4.627 *	14/14	58.25 [†]	13/14	<u>67.96</u>	6/8	70.04 [†]
GPT-4.1	Raw Model	89.82	-	59.45	-	2.131	-	52.69	-	60.91	-	61.10
	Role-Play	89.29	8/14	<u>63.47</u>	11/14 *	<u>4.270</u>	14/14	53.76	9/14	69.85	8/8	<u>72.35</u>
	CultureBank	90.80 *	11/14	60.00	9/14	4.226	14/14	56.76	14/14	68.11	8/8	72.04
	CAReDiO	<u>90.32</u>	10/14	63.54 *	11/14 *	4.336 [†]	14/14	<u>56.64</u> [†]	13/14	<u>69.66</u> [†]	8/8	73.37 ^{†‡}

Conclusion and Takeaways

(1) Motivated by culture theories, we identify the challenges of representativeness and distinctiveness in curating effective data for cultural alignment.

(2) We propose CArE DiO, an effective data optimization framework with two novel information-theoretic objectives.

(3) Using constructed data, we validate our method can achieve better cultural alignment across LLM

Value Compass Homepage: <https://valuecompass.github.io/>

