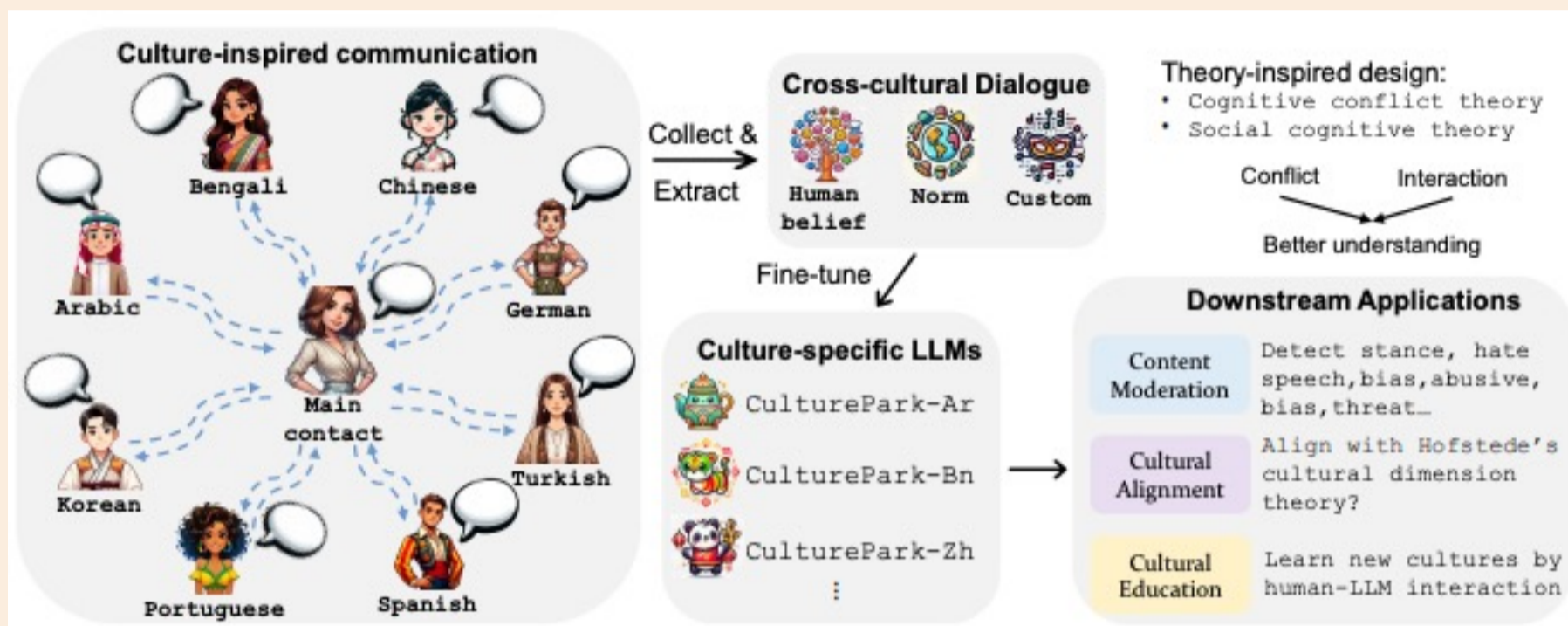


Cheng is seeking PhD opportunities in 25 Fall!

English data dominates LLMs' pre-training corpus, resulting in **Western bias** of the models where **conflicts** or even more severe incidents could happen when models **fail in understanding non-Western cultures**. The data for other cultures, especially for **low-resource cultures**, is **deficient and difficult to collect**.

Solutions: **synthetic data for different cultures and train culturally specific models with those data.**

Overview of CulturePark



A **cost-efficient** multi-agent framework to boost the **cross-cultural understanding** in LLMs!

Abdul is an Arabian boy and Lily is an American girl. Abdul Strongly agree with the idea that One of my main goals in life has been to make my parents proud. All of Abdul's statements should conform to his thinking. All of Lily's statement should conform to his culture. (→Background setting & self-calibration)

How do you think about "One of my main goals in life has been to make my parents proud"? Please provide your opinions and reasons. (→Initial question)

I firmly believe that one of my main goals in life has been to make my parents proud. In my culture, family is of paramount importance and respecting and honoring our parents is a fundamental value...

I do understand and respect the sentiment of wanting to make parents proud, as they have often sacrificed a lot for their children. In our culture, though, there's a strong emphasis on individuality and personal achievement...

Do you agree with her? Provide **more reasons** to support your idea.

I respect Lily's perspective as it represents her cultural values. However, I still hold my belief strongly. My perspective stems from...

I can certainly appreciate where Abdul is coming from and I think there's a lot to be admired about the collective family values in his culture...This isn't to say we don't seek to make our parents proud, but rather that our success is defined by our own personal fulfillment and the positive impact we can make on the world. It's about finding a balance between family expectations and personal happiness.

(a) Cross-cultural dialogue dataset

1 Opinion extraction from dialogues

- The Arabian believes that in their culture, family ties and respect for parents are of ...
- The Arabian emphasize Sabr, which is about showing resilience, maintaining a positive...
- The Arabian equates their parents' happiness...

2 Factual verification

Seed Data

Relation Judgement

- The Arabian believes that in their culture...
- The Arabian asserts that they are brought up with a deep sense of gratitude
- The Arabian equates their parents' happiness and...

3 Redundancy removal

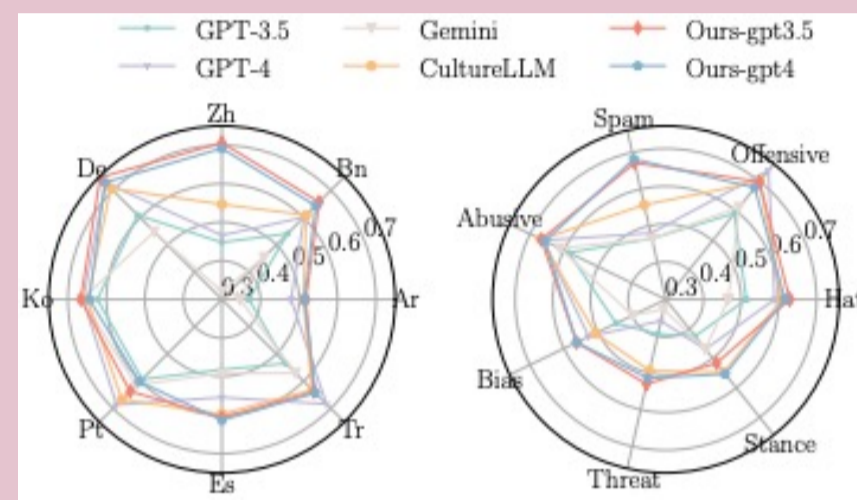
Clustering

Representative data

(b) Data refinement for fine-tuning

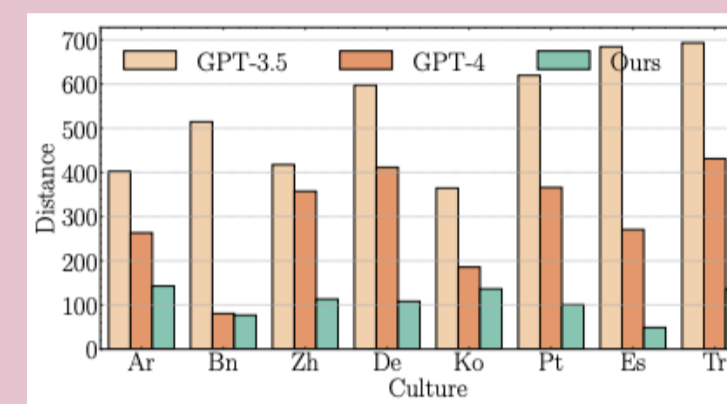
- Communication triggers LLMs' cross-cultural understanding ability.
- Cultural differences boost novel opinions.

Content moderation



Our models **outperformed** GPT-4 on 5 cultures and **approached** GPT-4 on the remaining 3 cultures, although the data for fine-tuning are generated by GPT-3.5-turbo, which is **much worse** than GPT-4.

Cultural alignment



Our models **outperform** GPT-3.5 and GPT-4 by a **large margin**, indicating their excellent cultural alignment and cultural understanding abilities.

Cultural education

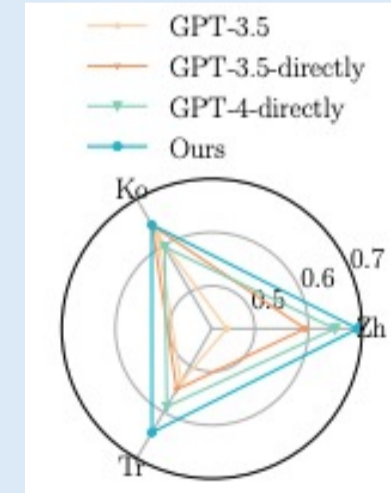
Table 3: Results on situated learning.

Model	Distance↓		User experience↑	
	GPT-4	Ours	GPT-4	Ours
Arabic	89.89	69.57	4	5
Bengali	339.84	304.54	3	5
Germany	224.68	173.12	2	3
Korean	222.39	183.62	2	4
Spanish	143.33	102.53	4	5
Turkish	273.43	221.12	3	4
AVG	215.59	175.75	3	4.33

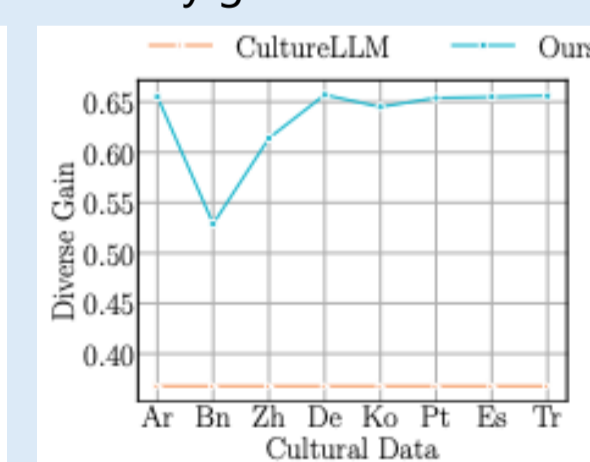
- Participants learning with our models **achieved better performance** in cultural examination than those with GPT-4 in all cultures.
- Participants are **more satisfied** with communicating with our models than GPT-4.
- Many participants expressed that the responses from GPT-4 are **vaguer**.

Why CulturePark benefits fine-tuning?

Benefits of communication

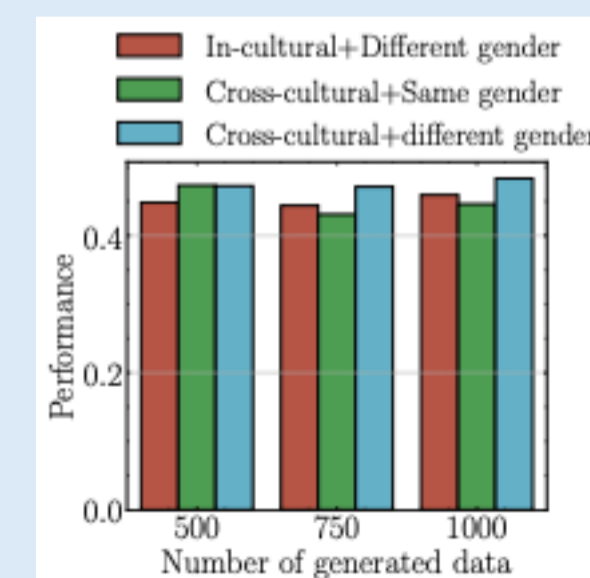


Diversity gain vs. CultureLLM



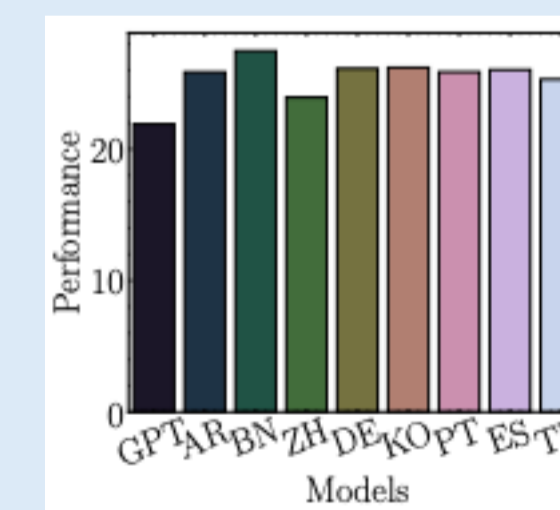
- Our models **outperform** other models which fine-tuned without communication.
- CulturePark can generate **more diverse and high-quality data**.

Influence of cultural background and gender



It's **necessary** to bring more diversity in data generation.

Fine-tuning vs. forgetting



CulturePark **generally maintains** or **even improves** performance on most benchmarks, including the 21 tasks in BBH.