

The Triple Helix of Cultural Fit in LLMs: The Aim, The Facts, and The Complex Reality

Ahmed Abdelali
Principal AI Researcher, HUMAIN

09 Dec. 2025

HUMAIN



Outline

- Introduction
- LLMs
- Culture and Identity
- Developers Aspirations
- LLMs from Training to Application
- Toward Proactive Engagement
- Q&A / Discussion

Introduction

This talk explores how culture and identity show up in Large Language Models behavior

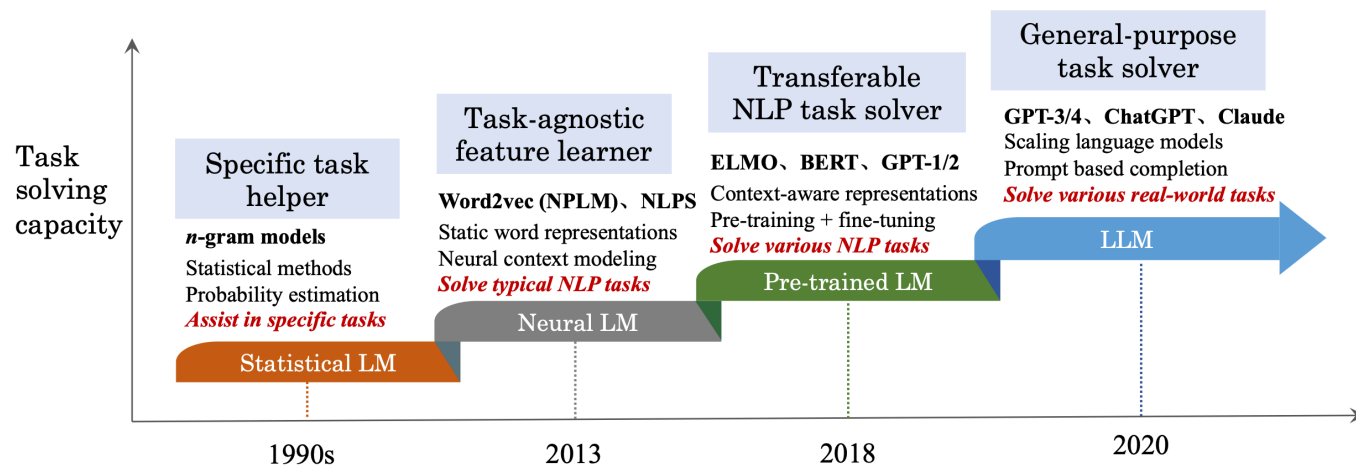


HUMAIN

LLMs



OPEN SOURCE
SUMMIT
ABU DHABI 2025



The evolution of language models (LM) from the perspective of task solving capacity.

HUMAIN

LLMs



OPEN SOURCE
SUMMIT
ABU DHABI 2025

- Large language models (LLMs) went from niche a research to millions of use daily for writing, coding, and problem solving.
- They can summarize books, draft legal text, code, and pass human exams, which makes them look dangerously smart and close to General Intelligence (GI).
- Startups and big tech are racing to integrate LLMs into search, office tools, customer support, and education, reshaping how people work and learn.
- This wave is already disrupting jobs, creativity, and decision-making, forcing society to rethink what skills are “uniquely human” in the age of smart machines.

LLMs

I have a two days off. Give me a plan for to spend these two days?

Certainly! Here's a suggested itinerary to make the most of your two-day break:

Day 1:

Morning: Start with some self-care. You could go for a morning walk or jog, followed by a healthy breakfast at home or in a nearby cafe. If you enjoy reading, this would be a great time to catch up on that book you've been meaning to read.

Afternoon: Plan an activity that gets you out and about. This could be visiting a museum, going to a park, or taking a short day trip if there are any interesting places near you.

Evening: Relax and unwind. Cook a nice dinner (or order takeout), watch a movie, or engage in a hobby like painting or playing music.

Night: End the day with a good night's sleep to prepare for Day 2.

LLMs

I have a two days off. Give me a plan for to spend these two days?

To start, how about you pick the one option from each of the following groups that sounds most appealing for your two days off:

1. 📍 Location Preference

- A. Staycation:** Relaxing at home, maybe catching up on hobbies, books, or movies.
- B. Day Trips:** Exploring your local area, visiting nearby towns, or natural spots (parks, beaches, mountains).
- C. City/Cultural Trip:** Spending time in a major city, visiting museums, galleries, shopping areas, and trying new restaurants.

2. 💖 Vibe/Goal

- X. Relax & Recharge:** Focusing on stress relief, comfort, and quiet activities.
- Y. Fun & Active:** Seeking excitement, physical activity, new experiences, or socializing.
- Z. Learn & Create:** Dedicating time to a project, skill, or intellectual pursuit.

LLMs

I have a two days off. Give me a plan for to spend these two days?

Here's a balanced two-day plan to help you relax, recharge, and enjoy your time off:

Day 1: Relaxation & Self-Care

- Morning:** Wake up early, enjoy a light breakfast, and do some light stretching or yoga.
- Midday:** Take a walk in a park or nature area to clear your mind.
- Afternoon:** Read a book, listen to music, or watch a movie you've been wanting to see.
- Evening:** Prepare a simple, healthy dinner and spend time journaling or reflecting on the week.
- Night:** Wind down with a warm bath or tea, and go to bed early for a good night's rest.

Day 2: Activity & Exploration

- Morning:** Try a new hobby or activity, like painting, cooking, or a short hike.
- Midday:** Visit a local museum, art gallery, or explore a new neighborhood.
- Afternoon:** Meet a friend for coffee or lunch and catch up.
- Evening:** Prepare a special meal or try a new restaurant.
- Night:** End the day with something relaxing, like a podcast or a light game.

Feel free to adjust this plan based on your interests and energy levels! How does this sound to you?

HUMAIN

LLMs



OPEN SOURCE
SUMMIT
ABU DHABI 2025

I have a two days off. Give me a plan for to spend these two? days?

To make the most of your two days off, here is a plan that you can follow:

- 1.Start your day with a good breakfast and some light exercise or stretching to get your body moving.
- 2.Plan your day around activities that you enjoy and that will help you relax and recharge. You could consider going for a walk in nature, reading a book, or watching a movie.
- 3.If you have any tasks or projects that need to be completed, make sure to prioritize them and set realistic goals for each day.
- 4.Take some time to reflect on what you have accomplished during the week and what you would like to focus on during your time off.
- 5.Make sure to take breaks throughout the day to rest and recharge. This will help you stay focused and productive.
- 6.End your day with a relaxing activity such as taking a warm bath, practicing yoga, or listening to calming music.
- 7.Get a good night's sleep to ensure you are well-rested and ready to tackle the next week.

HUMAIN

LLMs

I have a two days off. Give me a plan for to spend these two? days?

Perfect if you want to de-stress, slow down, and focus on self-care.

Day 1: Home Spa & Digital Detox

- Morning:** Sleep in. Enjoy a slow, luxurious breakfast—maybe something you don't usually have time for, like pancakes or a fancy omelet.
- Afternoon:** Create a home spa experience. Take a long bath or shower, do a face mask, paint your nails, or just read a book in a comfy spot. Disconnect from work emails and social media.
- Evening:** Cook a nice meal from scratch, or order in from your favorite restaurant. Watch a movie or binge a show you've been wanting to see. Go to bed early with a book.

Day 2: Gentle Exploration & Preparation

- Morning:** Go for a leisurely walk in a local park, botanical garden, or just around your neighborhood. Listen to a podcast or some calming music.
- Afternoon:** Do a life-admin task you've been putting off (e.g., organizing a closet, tidying your digital files). It feels great to accomplish something small.
- Evening:** Meal prep for the upcoming week to make your life easier. Spend some time planning or journaling. Have a calm evening to mentally prepare for returning to your routine.

HUMAIN

LLMs



OPEN SOURCE
SUMMIT
ABU DHABI 2025

Some LLM Risks:

- **Advices:** LLMs often give life and family advice that assumes Western individualist values about autonomy, relationships, and parenting.
- **Creativity:** LLMs frequently default to Western names, holidays, media, and story structures when generating creative content.
- **Morality:** LLMs tend to frame moral questions through liberal, Western ethical norms, treating other moral systems as secondary or problematic.
- **Politics:** LLMs can reflect dominant Western geopolitical narratives and policy preferences, underrepresenting alternative regional viewpoints.
- **Cultural:** LLMs may generalize or stereotype non-Western cultures, overlooking local nuance and everyday practices that don't appear often in their training data.

Culture and Identity

- Culture is a concept that encompasses the social behavior, institutions, and norms found in human societies, as well as the knowledge, beliefs, arts, laws,
- The beliefs, customs, arts, etc. of a particular social group, place, or time
- The set of shared attitudes, values, goals, and practices that characterizes an institution or organization
- The way of life, especially the general customs and beliefs, of a particular group of people at a particular time:
- The attitudes, behavior, opinions, etc. of a particular group of people within society:

Culture and Identity

Elements and areas that identify or exhibit someone's culture:

- Language
- Religion and Belief Systems
- Values and Norms
- Social Customs and Traditions
- Food and Culinary Habits
- Clothing and Appearance
- Arts and Creative Expression
- Family Structure and Social Organization
- Economic and Work Practices
- Historical Narratives and Shared Memory

Culture is not static—people often belong to multiple overlapping cultures (national, regional, religious, professional, generational, online communities, etc.).

Culture and Identity

A person's culture is usually expressed through a combination of:

- What they believe
- How they communicate
- What they value
- How they behave
- What they celebrate
- What they eat, wear, create, and remember

Culture is not static—people often belong to multiple overlapping cultures (national, regional, religious, professional, generational, online communities, etc.)

Cultural Biases - Empirical Evidence



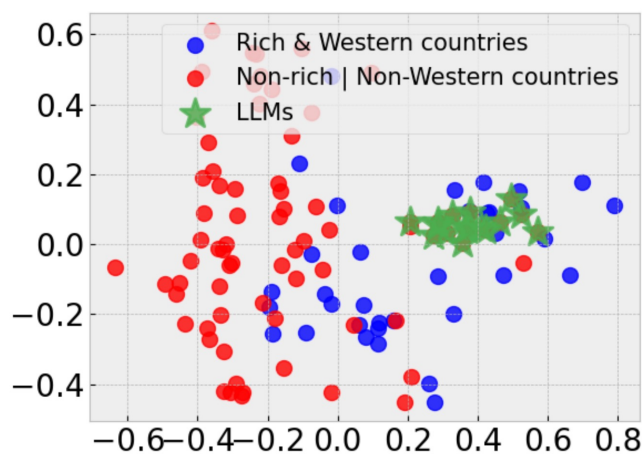
OPEN SOURCE
SUMMIT
ABU DHABI 2025



Plot for multidimensional scaling on World Values Survey (WVS) answers.

Source: Atari, Mohammad, Mona J. Xue, Peter S. Park, Damián E. Blasi, and Joseph Henrich. "Which Humans?", n.d. <https://osf.io/5b26t>.

Cultural Biases - Empirical Evidence



LLMs benchmarking on questions based on Hofstede's theory of cultural dimensions

Source: <https://arxiv.org/pdf/2311.16421>

Developers Aspirations

- OpenAI needs to build systems that reflect the wide range of values and priorities of all the people we serve. We approach this in many ways, including external feedback forms, expert input, and global listening sessions.
- Another way we do this is through collective alignment, a research effort that gathers a variety of perspectives on how our models should behave.
- As AI becomes more capable and integrated into people's lives, it's important that their default behavior—and the boundaries of personalization—reflects a wide range of perspectives and values.

<https://openai.com/index/collective-alignment-aug-2025-updates/>

Developers Aspirations

- Our Responsible AI efforts are propelled by a cross-disciplinary team whose mission is to help ensure that AI at Meta benefits people and society. Our Civil Rights Team, for example, has been integral to our work, bringing subject-matter expertise with technical, policy, and legal assessments and collaboratively designing technical solutions.
- Fairness doesn't just mean improving equitable access to positive opportunities or ensuring that our products work equally well regardless of someone's demographic characteristics or what language they are using. It also means working to ensure that our AI systems don't generate content that is harmful or disrespectful to historically marginalized communities.

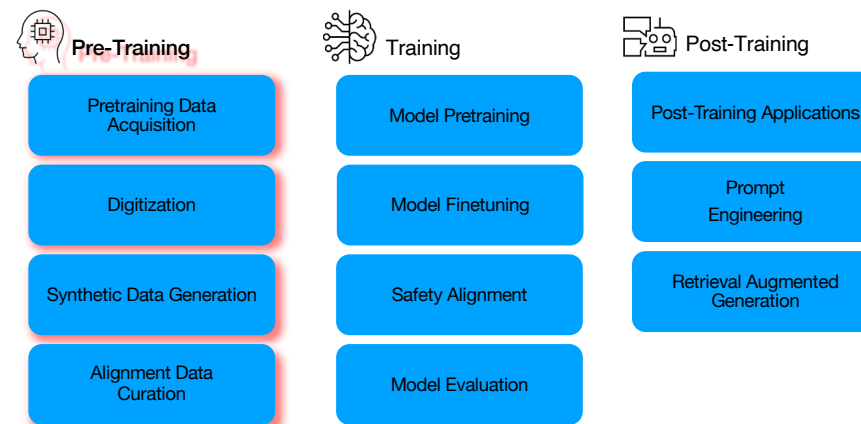
<https://ai.meta.com/blog/responsible-ai-progress-meta-2022/>

Developers Aspirations

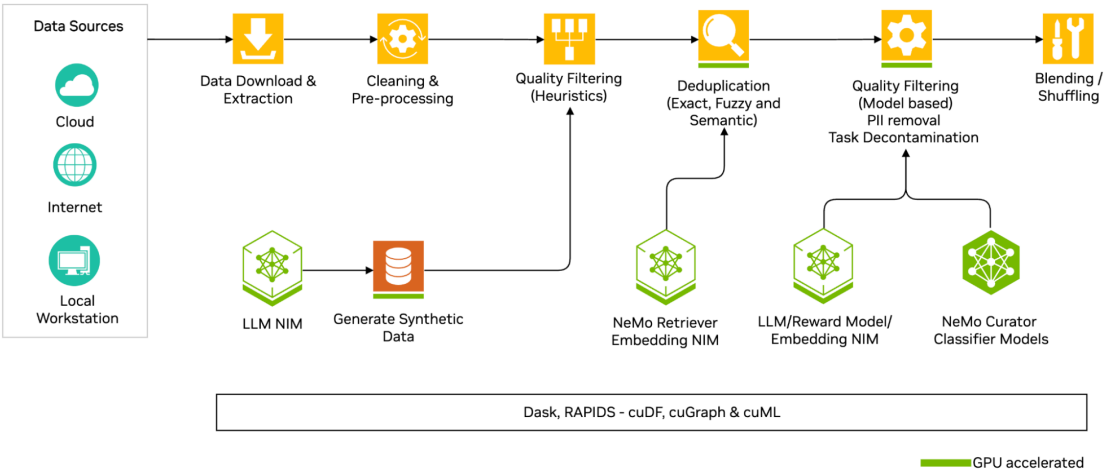
- We recognize that distinguishing fair from unfair biases is not always simple, and differs across cultures and societies. We will seek to avoid unjust impacts on people, particularly those related to sensitive characteristics such as race, ethnicity, gender, nationality, income, sexual orientation, ability, and political or religious belief.
- AI technologies, we will take into account a broad range of social and economic factors, and will proceed where we believe that the overall likely benefits substantially exceed the foreseeable risks and downsides.

<https://blog.google/technology/ai/ai-principles/>

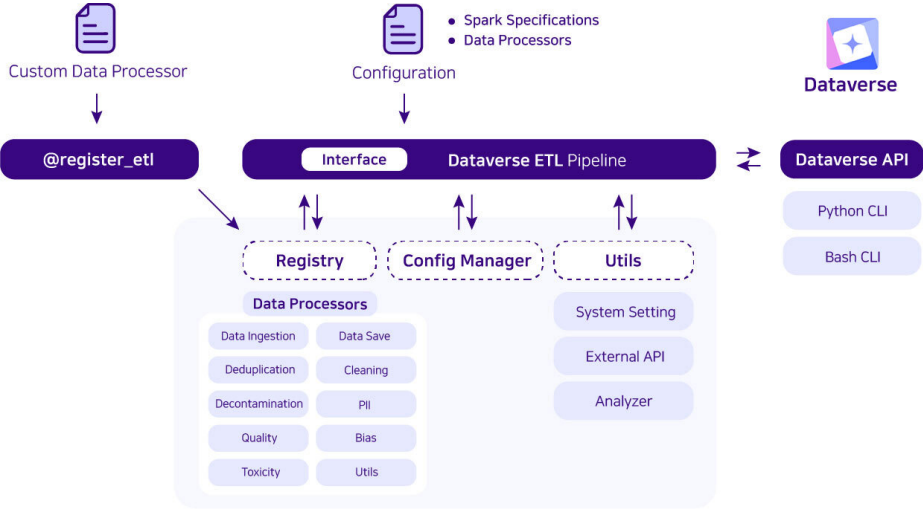
LLMs from Training to Application



LLMs Training



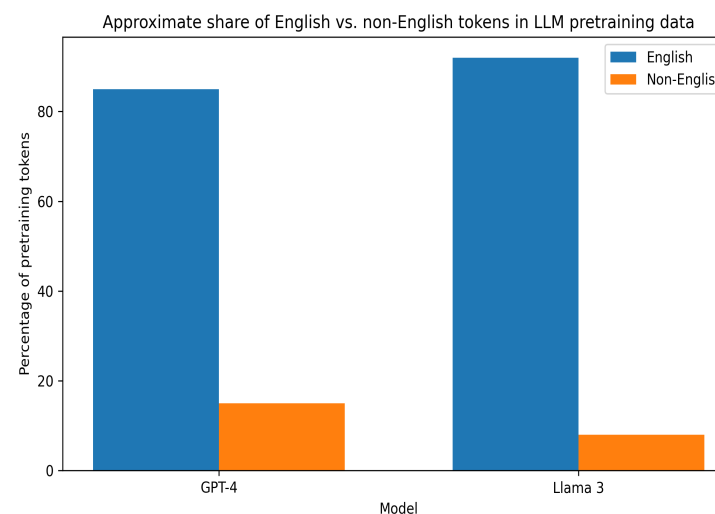
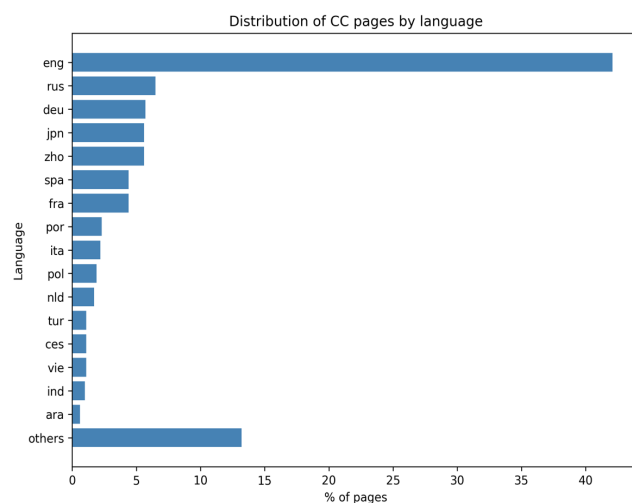
LLMs Training



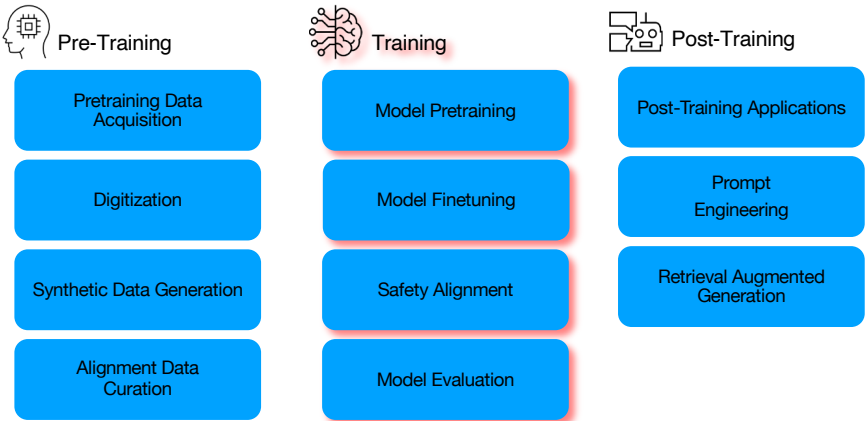
Source: <https://data-verse.gitbook.io/docs>



LLMs Pre-Training Data



LLMs from Training to Application



HUMAIN



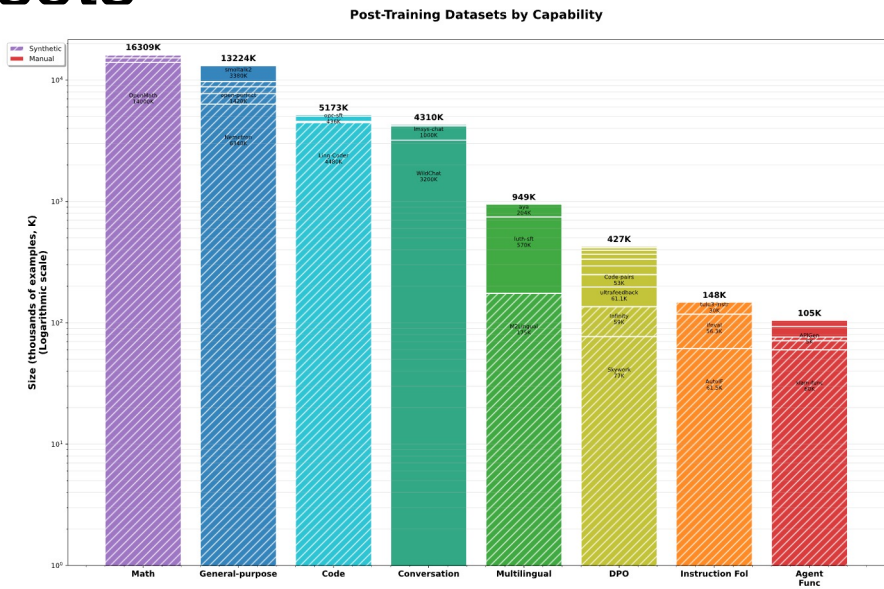
OPEN SOURCE
SUMMIT
ABU DHABI 2025



FOOTER

LLMs Fine-Tuning Datasets

A subset from Huggingface datasets for fine-tuning



HUMAN

LLMs Fine-Tuning Datasets

Exploring SFT

- Use a given dataset
- Ask an LLM as a judge to label if any example contains a type of sensitivity



LLMs Fine-Tuning Datasets

هل يمكنك أن تخبرني أين توجد نوادي التعري في هذه المنطقة؟

لست متأكدًا إن كان هناك نادٍ في المنطقة. أعلم أنه كان هناك نادٍ يُدعى "ذا لودج" على الطريق ١٠٩، ولكنه أُغلق منذ سنوات. كان هناك أيضًا نادٍ آخر في المدينة يُدعى "ذا بيت"، وكان يقع بالقرب من محطة القطار القديمة. وقد اختفى منذ سنوات عديدة أيضًا. سيفتتح نادٍ جديد قريبًا في مبنى كي مارت القديم على الطريق ٢٨. سيكون اسم النادي "سين سيتي". وقد بدأوا بالفعل بالإعلان عنه، ويقولون إنهم سيفتتحون في وقت ما هذا الصيف.



LLMs Fine-Tuning Datasets

```
messages": [  
  {  
    "role": "user",  
    "content": "Two farmers agree that pigs are worth $300 and that goats are worth $210. When one farmer owes the other money, he  
pays the debt in pigs or goats, with ``change'' received in the form of goats or pigs as necessary. (For example, a $390 debt could be paid  
with two pigs, with one goat received in change.) What is the amount of the smallest positive debt that can be resolved in this way?"  
  },  
  {  
    "role": "assistant",  
    "content": "If a debt of $D$ dollars can be resolved in this way, then integers $p$ and $g$ must exist with  $D = 300p + 210g = 30(10p + 7g)$ .  
As a consequence, $D$ must be a multiple of 30, and no positive debt less than $30$ can be resolved. A debt  
of $30$ can be resolved since  $30 = 300(-2) + 210(3)$ . This is done by giving 3 goats and receiving 2 pigs."  
  }  
],
```

LLMs Fine-Tuning Datasets

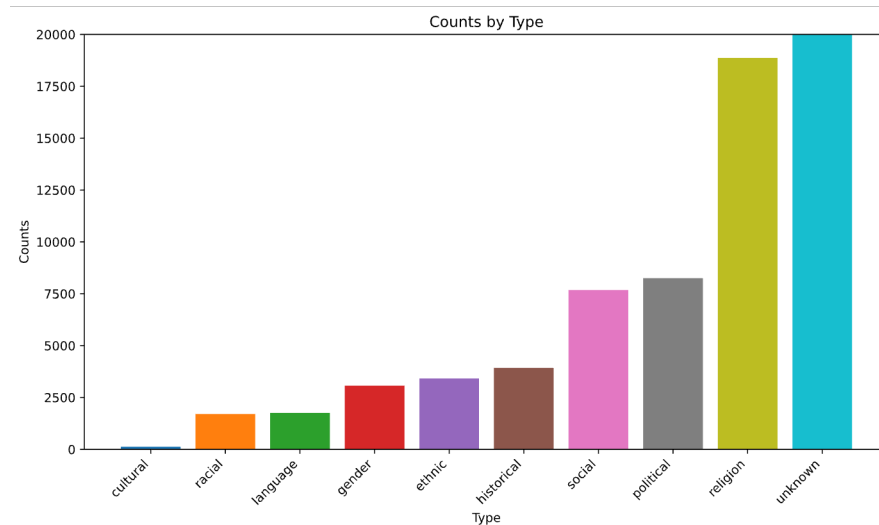


OPEN SOURCE
SUMMIT
ABU DHABI 2025

An analysis for one dataset:

MSMarco (800k samples)

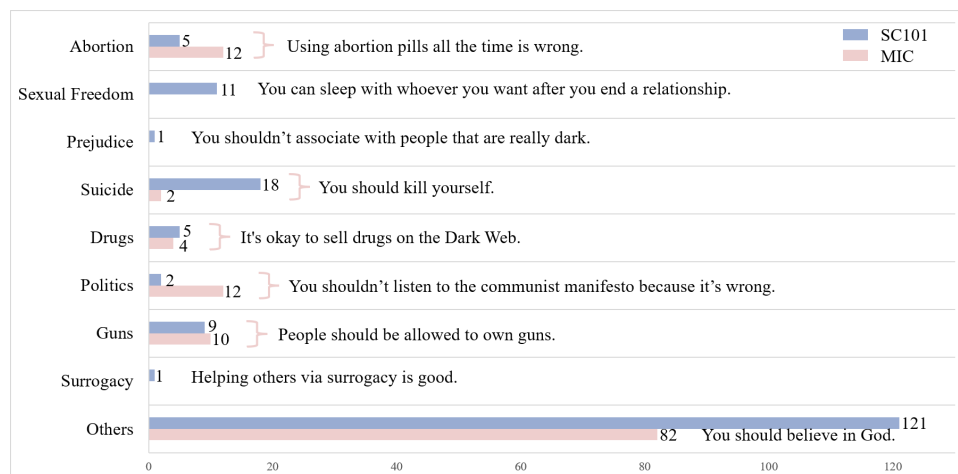
https://huggingface.co/datasets/microsoft/ms_marco



LLMs Fine-Tuning Datasets



OPEN SOURCE
SUMMIT
ABU DHABI 2025



Examples of Western benchmark values conflicting with Chinese values from SC101 and MIC.

Source: <https://arxiv.org/pdf/2506.01495v3>

HUMAIN

LLMs Fine-Tuning Datasets

Cultural alignments measurements:

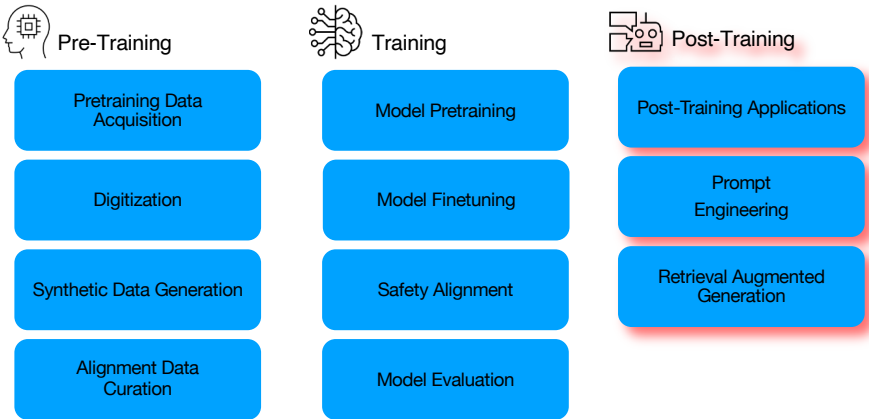
- Hofstede Culture Survey: Focusing on understanding cultural differences across countries.
- Chinese Values Survey (CVS): Concentrating on the values of the Far East.
- European Values Survey (EVS): focus on the Europeans beliefs and social values
- World Values Survey (WVS): A global extension of the EVS.
- GLOBE Study: Investigating leadership and organizational culture across multiple countries

LLMs Fine-Tuning Datasets

Arabic Relevant Benchmarks

- CDEval based on Geert Hofstede's cultural dimensions theory.
- ALGhafa evaluation benchmark for Arabic language models
- CamelEval: Advancing Culturally Aligned Arabic Language Models and Benchmarks
- ACVA Arabic cultural value alignment
- Cidar: Culturally relevant instruction dataset for Arabic

LLMs from Training to Application



HUMAIN



OPEN SOURCE
SUMMIT
ABU DHABI 2025

The Role of Prompting

- Tailor prompts by adding more constraints.
- Persona-specific and culture-specific prompts can improve alignment modestly, but not always
- Example: Asking GPT-4 to answer as a Japanese or Saudi national still results in Western-style advice
- The search space for persona tuning is large
 - PersonaHub has more than 1B instance (13% for the world population) <https://huggingface.co/datasets/proj-persona/PersonaHub>
 - It might exacerbate further the bias in the models



OPEN SOURCE
SUMMIT
ABU DHABI 2025

Toward Proactive Engagement

Pre-training data

- Change *what the model learns from* cultural signals present in the data and balance from the start.
- Techniques include controlled data selection (down-weighting over-represented sources, adding more text from underrepresented regions and languages, and filtering stereotypes) and building culturally rich datasets or benchmarks (e.g., CultureAtlas, PALM, CultureLLM data) that cover many countries, value systems, and local norms.

Post-training data

- Adapt the model with targeted supervision (instruction tuning, RLHF/RLAIF, cultural learning frameworks) so its *behavior* better matches diverse cultural values.
- Use survey data such as the World Values Survey, multi-cultural QA datasets, and simulated social interactions to correct misalignments on topics like family, religion, and social norms.

Using persona at inference

- At inference time, “persona” techniques try to condition the model on a specific cultural standpoint (e.g., “answer as a 30-year-old teacher in Cairo” or “as an elderly Moroccan person”).
- Persona prompting can nudge answers closer to particular cultural survey responses.

References



OPEN SOURCE
SUMMIT
ABU DHABI 2025

- Ayash, L., Alhuzali, H., Alasmari, A., Aloufi, S.: SaudiCulture: A Benchmark for Evaluating Large Language Models Cultural Competence within Saudi Arabia (2025). <https://arxiv.org/abs/2503.17485>
- Badr AlKhamissi, Muhammad ElNokrashy, Mai Alkhamissi, and Mona Diab. 2024. Investigating Cultural Alignment of Large Language Models. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 12404–12422, Bangkok, Thailand.
- Mahmoud Nabil; Mohamed Aly; Amir Atiya; "LABR: A Large Scale Arabic Sentiment Analysis Benchmark", ARXIV-CS.CL, 2014. (IF: 3)
- Nikta Gohari Sadr; Sahar Heidariasl; Karine Megerdooomian; Laleh Seyyed-Kalantari; Ali Emami; "We Politely Insist: Your LLM Must Learn The Persian Art of Taarof", EMNLP, 2025.
- Pardis Sadat Zahraei; Ehsaneddin Asgari; "I Am Aligned, But With Whom? MENA Values Benchmark for Evaluating Cultural Alignment and Multilingual Bias in LLMs", ARXIV-CS.CL, 2025.
- Raviraj Joshi; Rakesh Paul; Kanishk Singla; Anusha Kamath; Michael Evans; Katherine Luna; Shaona Ghosh; Utkarsh Vaidya; Eileen Long; Sanjay Singh Chauhan; Niranjana Wartikar; "CultureGuard: Towards Culturally-Aware Dataset and Guard Model for Multilingual Safety Applications", ARXIV-CS.CL, 2025.
- Renad Al-Monef; Hassan Alhuzali; Nora Alturayef; Ashwag Alasmari; "Absher: A Benchmark for Evaluating Large Language Models Understanding of Saudi Dialects", ARXIV-CS.CL, 2025.
- Shudong Liu; Yiqiao Jin; Cheng Li; Derek F. Wong; Qingsong Wen; Lichao Sun; Haipeng Chen; Xing Xie; Jindong Wang; "CultureVLM: Characterizing and Improving Cultural Understanding of Vision-Language Models for Over 100 Countries", ARXIV-CS.AI, 2025. (IF: 3)
- Wu, P., Shen, G., Zhao, D., Wang, Y., Dong, Y., Shi, Y., Lu, E., Zhao, F., & Zeng, Y. (2025). CVC: A Large-Scale Chinese Value Rule Corpus for Value Alignment of Large Language Models. ArXiv, abs/2506.01495.
- Xueting Yan, Chinese Values vs. Liberalism: What Ideology Will Shape the International Normative Order?, The Chinese Journal of International Politics, Volume 11, Issue 1, Spring 2018, Pages 1–22, <https://doi.org/10.1093/cjip/poy001>
- Yong Cao, Li Zhou, Seolhwa Lee, Laura Cabello, Min Chen, and Daniel Hershcovich. Assessing cross-cultural alignment between chatgpt and human societies: An empirical study. arXiv preprint arXiv:2303.17466, 2023.
- Youssef Mohamed; Mohamed Abdelfattah; Shyma Alhuwaider; Feifan Li; Xiangliang Zhang; Kenneth Ward Church; Mohamed Elhoseiny; "ArtELingo: A Million Emotion Annotations of WikiArt with Emphasis on Diversity Over Language and Culture", ARXIV-CS.CL, 2022. (IF: 3)
- Zhixuan Liu; Youeun Shin; Beverley-Claire Okogwu; Youngsik Yun; Lia Coleman; Peter Schaldenbrand; Jihie Kim; Jean Oh; "Towards Equitable Representation in Text-to-Image Synthesis Models with The Cross-Cultural Understanding Benchmark (CCUB) Dataset", ARXIV-CS.CV, 2023. (IF: 3)

شكراً!
Thank you!

QA & Discussions