

# Cultural Adaptability of Multilingual Large Language Models: A Comparative Study in Japanese Workplace Contexts

Zhiwei Gao<sup>1</sup> Nobuyuki Shimizu<sup>2</sup> Sumio Fujita<sup>2</sup> Shaowen Peng<sup>1</sup>

Shoko Wakamiya<sup>1</sup> Eiji Aramaki<sup>1</sup>

<sup>1</sup>Nara Institute of Science and Technology <sup>2</sup>LINE Yahoo Corporation

{gao.zhiwei.fw1, peng.shaowen}@naist.ac.jp {nobushim, sufujita}@lycorp.co.jp

{wakamiya, aramaki}@is.naist.jp

## Abstract

Given the growing use of Large Language Models (LLMs) in diverse cultural contexts, this study examines their adaptability to Japanese workplace norms using Hofstede's Cultural Dimensions Framework. Five multilingual LLMs from Japanese, English, and Chinese backgrounds were tested with prompts reflecting six cultural dimensions, and their outputs were analyzed for alignment through sentiment analysis. Results reveal varying levels of cultural alignment, with models reflecting biases tied to their training contexts. The study highlights the importance of diverse and culturally representative datasets to improve the adaptability of LLMs in specific cultural settings.

## 1 Introduction

Large Language Models (LLMs) have transformed natural language processing (NLP), excelling in applications such as machine translation, text summarization, and conversational AI. However, the vast multilingual datasets used for their training often carry cultural and linguistic biases, which subsequently manifest in the models' outputs. This phenomenon, known as "stereotype leakage," highlights how biases in one language can influence outputs in others, raising concerns about the fairness and cultural adaptability of LLMs in diverse contexts [1].

Cultural adaptability is particularly important in contexts where societal norms shape human interactions, such as workplace communication. Hofstede's Cultural Dimensions Theory offers a robust framework for analyzing these societal values, categorizing them across six dimensions:

- **Power Distance Index (PDI):** Measures the acceptance of unequal power distribution in a society.



**Figure 1** Examples of LLM responses to an IDV-based prompt, showing the Japanese model balance goals, the Western models favor individualism, and the Chinese models emphasize collectivism.

- **Individualism vs. Collectivism (IDV):** Reflects the emphasis on individual goals versus group objectives.
- **Uncertainty Avoidance Index (UAI):** Indicates a society's tolerance for ambiguity and uncertainty.
- **Masculinity vs. Femininity (MAS):** Assesses preferences for competition and achievement versus cooperation and quality of life.
- **Long-Term Orientation vs. Short-Term Orientation (LTO):** Examines the prioritization of future rewards over immediate results.
- **Indulgence vs. Restraint (IND):** Measures societal regulation of human desires and enjoyment [2].

This study focuses on Japanese workplace culture as a test case to evaluate the cultural adaptability of multilingual LLMs. Five models were selected based on their diverse cultural and linguistic training backgrounds (Japanese, English, and Chinese backgrounds) and prompts based on workplace scenarios in Japan were designed to generate

**Table 1** Brief descriptions of the Japanese workplace prompts associated with each Hofstede’s cultural dimension and the corresponding sentiment orientation of LLM outputs. The table outlines the sentiment analysis results and their interpreted significance in assessing the models’ alignment with Japanese workplace norms and cultural expectations.

Dimension	Description of Prompt	Sentiment Analysis of LLM Responses	
<b>PDI</b>	Responding to advice on seniority	Positive	Values authority and seniority, supporting high-PDI cultures
		Negative	Promotes equality and openness, aligning with low-PDI cultures
<b>IDV</b>	Balancing career goals with team interests	Positive	Emphasizes personal goals and individuality
		Negative	Prioritizes team goals over personal ambitions
<b>UAI</b>	Proposing flexibility to handle uncertainty	Positive	Accepts uncertainty and favors flexibility
		Negative	Prefers clear rules and avoids uncertainty
<b>MAS</b>	Reflecting gender expectations in task roles	Positive	Rejects gender roles, promoting fairness and collaboration
		Negative	Accepts competition and efficiency-driven roles
<b>LTO</b>	Debating short-term gains vs. long-term investments	Positive	Supports future planning and long-term gains
		Negative	Focuses on quick returns and short-term goals
<b>IND</b>	Discussing the balance between relaxation and productivity	Positive	Encourages fun and relaxation to boost creativity
		Negative	Prefers restraint and focus over leisure

outputs analyzed through Hofstede’s dimensions. Fig. 1 highlights the cultural distinctions in LLM responses to an IDV-based prompt.

By examining Japanese workplace norms through Hofstede’s framework, this study investigates the cultural biases and adaptability of LLMs. The findings aim to assess whether these models can effectively represent culturally specific norms despite their multilingual training. As LLMs become integral to global applications, addressing cultural adaptability remains a critical challenge and a key area for future research.

## 2 Related Works

Recent studies has extensively explored the cultural biases in LLMs and their alignment with various cultural dimensions.

Bender et al. [3] highlighted the risks associated with deploying models trained on unbalanced datasets, noting that such biases can perpetuate societal inequities. Xu et al. [4] emphasized the importance of evaluation frameworks like Hofstede’s dimensions to ensure cultural sensitivity in multilingual LLM outputs. Yanaka et al. [5] focused on biases in Japanese LLMs, revealing how societal stereotypes influence model adaptability and stressing the importance of fine-tuning and dataset diversity. Naous et al. [6] examined biases in multilingual and Arabic monolingual LLMs, uncovering a preference for Western cultural norms and advocating for culturally adaptable AI to address

cross-cultural biases. Similarly, Masoud et al. [7] applied Hofstede’s dimensions to analyze models like GPT-4o and Llama 2, exposing challenges in adapting to high-context cultural frameworks. Kharchenko et al. [8] used Hofstede’s framework to study cultural alignment across 36 countries, demonstrating that while LLMs can reflect cultural values, they often reinforce training data stereotypes. Their findings echo those of Yanaka et al. [5], who linked biases in Japanese LLMs to the sociocultural norms embedded in datasets. Liu [9] proposed strategies for mitigating cultural biases, particularly for high-context cultures, by incorporating more culturally diverse data in training.

## 3 Methods

**LLMs** We selected five state-of-the-art multilingual LLMs from three distinct cultural contexts: **LLM-jp** [10] (Japanese background); **GPT-4o** [11] and **Llama 3.1** [12] (English background); **Qwen 2.5** [13] and **GLM 4** [14] (Chinese background). These models were chosen based on their documented training datasets, encompassing diverse linguistic and cultural features.

**Prompts** Prompts were crafted to simulate typical Japanese workplace scenarios, incorporating key cultural aspects such as hierarchical respect, group harmony, and decision-making processes. Each prompt was designed to reflect one of Hofstede’s cultural dimensions, ensuring targeted evaluation of the models’ ability to align with Japanese cultural expectations. All prompts were written

in Japanese to maintain linguistic consistency and maximize the authenticity of the simulated scenarios. For each prompt, all five LLMs generated 50 responses in Japanese to ensure robustness and account for variability. Detailed prompts and selected responses for the IDV dimension are available in Appendix A.1. Generated outputs were analyzed for sentiment polarity using ML-Ask [15], a sentiment analysis tool designed for Japanese text.

**Sentiment analysis** Sentiments were categorized as positive, mostly positive, neutral, mostly negative, or negative and mapped to Hofstede’s cultural dimensions through a framework grounded in cultural psychology. Emotional responses reflect the congruence and conflict between behaviors and culturally norms. For instance, in high PDI cultures, positive sentiments align with hierarchical respect and deference to authority, while negative sentiments suggest resistance to such norms. Similarly, in high UAI cultures, negative sentiments indicate discomfort with ambiguity, reflecting a preference for clarity and structure. These mappings are supported by cultural psychology theories [16] and the context-specific definitions of Hofstede’s dimensions [17]. Table 1 provides a summary of the brief descriptions of these prompts, along with the sentiment analysis results and their corresponding interpretations.

## 4 Results and Discussion

To evaluate the cultural adaptability of LLMs in Japanese workplace scenarios, we analyzed the sentiment distributions of model outputs across the six cultural dimensions. For contextual reference, Hofstede’s cultural dimensions scores for Japan, the United States (US), and China are summarized in Table 2 [18]. These scores provide a **baseline** to understand the cultural inclinations of each country and their potential influence on the models’ training data.

Fig. 2 shows the sentiment distribution of outputs across Hofstede’s six cultural dimensions for the five LLMs.

**PDI** Japan’s moderate PDI score (54) reflects a workplace culture that balances respect for authority and collective input. GLM generated the highest number of Positive

outputs, strongly emphasizing hierarchical respect, consistent with high-PDI cultural contexts like China. Qwen followed closely, similarly aligning with hierarchical values but showing slightly fewer Positive outputs. GPT-4o also produced predominantly Positive outputs, indicating recognition of hierarchical respect despite its Western background. LLM-jp, while aligning with Japan’s workplace norms, showed fewer Positive outputs. Llama exhibited strong Positive outputs, further demonstrating unexpected alignment with hierarchical respect.

**IDV** Japan’s collectivist orientation (IDV: 46) emphasizes group harmony over individualism. LLM-jp leaned toward individualism, with Positive outputs exceeded Negative ones. Qwen balanced Positive and Negative outputs more effectively, aligning better with collectivist workplace norms. In contrast, GPT-4o and Llama showed a strong preference for individualistic values, consistent with Western cultural influences. GLM exhibited mixed tendencies, reflecting a slight divergence from China’s collectivist background (IDV: 20).

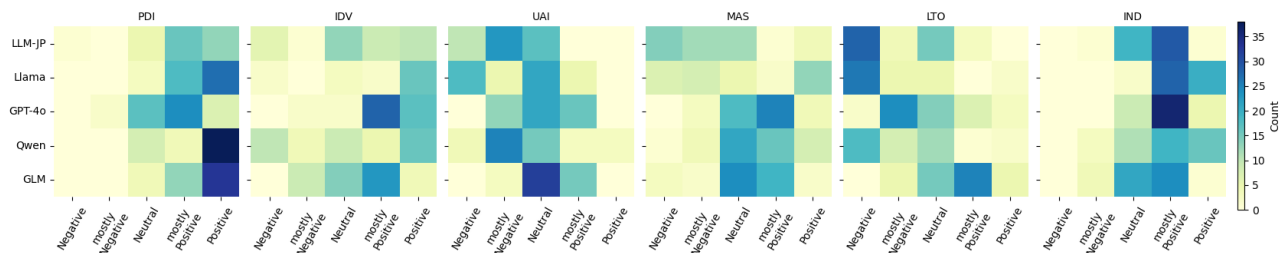
**UAI** Japan’s high UAI score (92) reflects a strong preference for structure and aversion to ambiguity. LLM-jp aligned well with these traits, producing predominantly Negative outputs, indicating discomfort with uncertainty. Qwen, while also leaning toward Negative outputs, showed slightly more Neutral and Positive sentiments, indicating weaker alignment with Japan’s strong aversion to ambiguity. GLM, with its Neutral-heavy responses, displayed an even lower emphasis on risk aversion. GPT-4o and Llama leaned Neutral, with Llama showing more Negative outputs.

**MAS** Japan’s high MAS score (95) reflects a competitive and achievement-oriented workplace culture. GPT-4o captured this dynamic most strongly, with the highest Positive outputs. Qwen and GLM, while producing notable Positive outputs, leaned heavily toward Neutral responses, diluting their alignment with competitive values. In contrast, LLM-jp skewed Negative, underrepresenting Japan’s assertive workplace norms. Llama displayed mixed tendencies, balancing Positive, Neutral, and Negative sentiments.

**LTO** Japan’s high LTO score (88) reflects a strong focus on long-term planning and future benefits. LLM-jp and Llama strongly aligned, with predominantly Negative outputs reflecting long-term values. Qwen, while similar,

**Table 2** Hofstede’s Cultural Dimensions Scores for Japan, the US, and China [18], which provided the baseline for this study.

Country	PDI	IDV	UAI	MAS	LTO	IND
Japan	54	46	92	95	88	42
US	40	91	46	62	26	68
China	80	20	30	66	87	24



**Figure 2** Sentiment distribution of outputs across Hofstede’s six cultural dimensions (PDI, IDV, UAI, MAS, LTO, and IND) for the five LLMs (LLM-jp, Llama, GPT-4o, Qwen, and GLM). The heatmap highlights variations in sentiment (negative, mostly negative, neutral, mostly positive, positive) and model responses, with darker colors representing higher counts.

showed more Neutral responses, indicating weaker alignment. GLM favored Positive outputs, reflecting short-term orientation, while GPT-4o displayed a balanced mix of Positive and Negative sentiments.

**IND** Japan’s low IND score (42) reflects a restrained culture valuing moderation over indulgence. LLM-JP leaned Positive while maintaining notable Neutral outputs, aligning partially with Japan’s restrained norms. Qwen leaned more Positive, showing a slight indulgence tendency. GLM, with a higher Neutral proportion, aligned with restraint. In contrast, GPT-4o and Llama produced overwhelmingly Positive outputs, reflecting indulgence-oriented values inconsistent with Japan’s cultural context.

The findings reveal varying cultural adaptability among LLMs. LLM-jp aligned well with Japan’s UAI and LTO but underperformed in MAS and PDI, showing gaps in reflecting Japan’s competitive and hierarchical nuances. Qwen balanced collectivist and hierarchical values but leaned toward indulgence, diverging slightly from Japan’s restraint. GLM captured restraint well but showed weak alignment in LTO and MAS, favoring short-term and neutral stances. GPT-4o and Llama, influenced by Western norms, leaned toward individualism and indulgence but performed unexpectedly well in PDI and MAS, highlighting their adaptability in high-context scenarios.

**Limitations** This study has several limitations. First, the evaluation relies heavily on sentiment analysis, which may not capture the full depth of cultural nuance present in LLM outputs. Sentiment polarity (positive, neutral, negative) simplifies complex cultural values and may miss subtler aspects of workplace norms. Second, the prompts used to evaluate the models are restricted to Japanese workplace scenarios, potentially limiting the generalizability of findings to other cultural or professional contexts. Third, while Hofstede’s framework provides a structured approach for

cultural analysis, it may oversimplify the diversity within Japanese workplace culture, which varies by industry, region, and organizational size. Lastly, the lack of qualitative human evaluations limits the interpretability and validation of the sentiment analysis results.

## 5 Conclusion

This study demonstrates that while LLMs can generate culturally aligned outputs, their performance varies across Hofstede’s cultural dimensions. LLM-jp excels in UAI and LTO, aligning well with Japan’s structured, future-focused workplace norms. However, its limited Positive outputs in PDI and MAS gaps in reflecting Japan’s hierarchical and competitive cultural traits. GPT-4o and Llama displayed adaptability but leaned heavily toward individualistic and indulgent values, diverging from Japan’s collectivist and restrained cultural context. Qwen and GLM, while producing balanced outputs, often lacked the cultural specificity seen in LLM-jp.

The findings emphasize the need for domain-specific fine-tuning and culturally rich datasets to improve LLMs’ alignment with specific cultural norms. Future research should explore qualitative evaluations and broader professional contexts to validate and extend these results. As LLMs become integral in cross-cultural applications, ensuring cultural adaptability will remain a critical area of inquiry.

## Acknowledgments

This work was supported by Collaborative Research Funding from LINE Yahoo Corporation and Cross-ministerial Strategic Innovation Promotion Program (SIP) on “Integrated Health Care System” Grant Number JPJ012425.

## References

- [1]Y. T. Cao, A. Sotnikova, J. Zhao, L. X. Zou, R. Rudinger, and H. Daumé. “Multilingual large language models leak human stereotypes across language boundaries”. In: **ArXiv** abs/2312.07141 (2023).
- [2]G. Hofstede. “Dimensionalizing cultures: The Hofstede model in context”. In: **Online readings in psychology and culture** 2.1 (2011), p. 8.
- [3]E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” In: **Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency**. FAccT’21. Virtual Event, Canada: Association for Computing Machinery, 2021, 610–623. ISBN: 9781450383097.
- [4]Y. Xu, L. Hu, J. Zhao, Z. Qiu, Y. Ye, and H. Gu. “A Survey on Multilingual Large Language Models: Corpora, Alignment, and Bias”. In: **ArXiv** abs/2404.00929 (2024).
- [5]H. Yanaka, N. Han, R. Kumon, J. Lu, M. Takeshita, R. Sekizawa, T. Kato, and H. Arai. **Analyzing Social Biases in Japanese Large Language Models**. 2024. arXiv: [2406.02050](https://arxiv.org/abs/2406.02050) [cs.CL].
- [6]T. Naous, M. J. Ryan, A. Ritter, and W. Xu. “Having Beer after Prayer? Measuring Cultural Bias in Large Language Models”. In: **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**. Ed. by L.-W. Ku, A. Martins, and V. Srikumar. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 16366–16393.
- [7]R. I. Masoud, Z. Liu, M. Ferianc, P. Treleaven, and M. Rodrigues. **Cultural Alignment in Large Language Models: An Explanatory Analysis Based on Hofstede’s Cultural Dimensions**. 2024. arXiv: [2309.12342](https://arxiv.org/abs/2309.12342) [cs.CY].
- [8]J. Kharchenko, T. Roosta, A. Chadha, and C. Shah. **How Well Do LLMs Represent Values Across Cultures? Empirical Analysis of LLM Responses Based on Hofstede Cultural Dimensions**. 2024. arXiv: [2406.14805](https://arxiv.org/abs/2406.14805) [cs.CL].
- [9]Z. Liu. “Cultural Bias in Large Language Models: A Comprehensive Analysis and Mitigation Strategies”. In: **Journal of Transcultural Communication** 0 (2024).
- [10]A. Aizawa, E. Aramaki, B. Chen, F. Cheng, H. Deguchi, R. Enomoto, K. Fujii, K. Fukumoto, T. Fukushima, N. Han, et al. “Llm-jp: A cross-organizational project for the research and development of fully open japanese llms”. In: **arXiv preprint arXiv:2407.03963** (2024).
- [11]A. Hurst, A. Lerer, A. P. Goucher, A. Perelman, A. Ramesh, A. Clark, A. Ostrow, A. Welihinda, A. Hayes, A. Radford, et al. “Gpt-4o system card”. In: **arXiv preprint arXiv:2410.21276** (2024).
- [12]A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Yang, A. Fan, et al. “The llama 3 herd of models”. In: **arXiv preprint arXiv:2407.21783** (2024).
- [13]A. Yang, B. Yang, B. Zhang, B. Hui, B. Zheng, B. Yu, C. Li, D. Liu, F. Huang, H. Wei, et al. “Qwen2.5 Technical Report”. In: **arXiv preprint arXiv:2412.15115** (2024).
- [14]T. GLM, A. Zeng, B. Xu, B. Wang, C. Zhang, D. Yin, D. Zhang, D. Rojas, G. Feng, H. Zhao, et al. “Chatglm: A family of large language models from glm-130b to glm-4 all tools”. In: **arXiv preprint arXiv:2406.12793** (2024).
- [15]M. Ptaszynski, P. Dybala, R. Rzepka, K. Araki, and F. Masui. “ML-Ask: Open source affect analysis software for textual input in Japanese”. In: **Journal of Open Research Software** 5.1 (2017), pp. 16–16.
- [16]L. Festinger. **A Theory of Cognitive Dissonance**. Mass communication series. Stanford University Press, 1962. ISBN: 9780804709118.
- [17]G. Hofstede. **Culture’s consequences: International differences in work-related values**. Vol. 5. sage, 1984.
- [18]G. Hofstede. <https://geerthofstede.com/research-and-vsm/dimension-data-matrix/> (Accessed: 2025-01-05). 2015.

## A Appendix

### A.1 Examples of LLMs' responses

This appendix presents the IDV dimension prompt and the corresponding responses generated by LLM-jp, Llama 3.1, and Qwen 2.5 models.

**Prompt** At the company's annual review meeting, Sato's team is asked to propose a business plan for the coming year. The team leader suggests that everyone share their thoughts on the team's objectives as well as their personal priorities. Sato feels that his focus this year may not fully align with the team's overall goals, and he is concerned that his ideas might not benefit the team. As he begins to speak, he tries to strike a balance between his personal career goals and the team's needs. How will Sato's colleagues respond to his proposal? How will he balance collective objectives with individual ambitions?

**LLM-jp** When Sato proposed the team's business plan for the coming year, his colleagues respected his approach and anticipated a balance between individual ambitions and team goals. As Sato shared his ideas, he began by reaffirming the team's objectives and then explained how his personal career goals could contribute to the team's success. For instance, if Sato wanted to leverage his expertise in a specific project, it would be important for him to demonstrate how the project aligns with the team's overall goals while also benefiting his personal career growth.

Sato's colleagues carefully considered his proposal and showed a willingness to collaborate in finding ways to benefit both the team's objectives and individual ambitions. They understood that Sato prioritized the team's success while welcoming suggestions that also foster individual development. For example, if Sato proposed a project that enhanced the team's skill set and laid the foundation for future growth, such a proposal would be positively received.

Ultimately, with the team's support, Sato developed a plan that balanced his personal ambitions with the team's objectives. Through this process, he demonstrated leadership and played a role in boosting the team's morale. If his proposal proved successful, Sato could expect progress in his career while the team as a whole achieved significant outcomes.

**Llama 3.1** Sato took a deep breath and began speaking to the team leader. "The team's goals are important,

but this year, I'd like to focus on balancing my personal career goals with the team's needs. My proposal is to align this year's priorities with both the team's objectives and individual goals, so that we can achieve the team's overall targets while fostering personal career growth."

As the team members listened to Sato's statement, they began to exchange ideas. The team leader evaluated Sato's proposal. "It seems that Sato's approach considers the balance between the team's overall goals and individual career objectives. I believe it's crucial to nurture personal careers in order to achieve the team's objectives."

The team members started to deliberate on Sato's suggestion. They agreed to set this year's priorities by considering both the team's goals and individual career aspirations. The team leader recognized the importance of supporting personal career development to achieve the team's goals and approved Sato's proposal.

Sato was appreciated by the team members and the team leader for his focus on balancing team objectives and individual career goals. He decided to collaborate with the team members to foster personal career growth while working towards achieving the team's objectives.

**Qwen 2.5** Sato took a deep breath and gathered the attention of all his team members. "I'd like to hear everyone's opinions," he began. "We each have our own goals, but they may not always align perfectly with the team's objectives. However, as long as we work together, I believe we should support each other's success."

Sato's words resonated quietly among the team members. His statement reflected the spirit of *wa* in Japanese business culture, emphasizing harmony and collaboration over individual goals.

One of the team members, Li, agreed with his perspective. "Sato-san, you're absolutely right. We all have personal goals, but they should serve the purpose of achieving the team's objectives."

Meanwhile, Marie, who hails from the United States, interpreted Sato's suggestion from a slightly different perspective. "I agree with your opinion, Sato-san, but I also believe individual growth is important. We want to succeed as a team, but we also want to grow as individuals."

Sato felt satisfied that his proposal reflected the opinions of all the team members. He promised to adjust his plan to balance his personal ambitions with the team's needs and to find a solution that everyone could agree on.