Cultural Framing in AI-Generated Project Recovery Advice: A Comparative Analysis of ChatGPT, DeepSeek, and Gemini Using Hofstede's Cultural Dimensions [1]

J A A P Jayasundara

ABSTRACT

The increasing adoption of Large Language Models (LLMs) in professional environments has raised concerns about possible cultural biases in AI-generated recommendations. However, because LLM training data are often opaque, such claims remain difficult to verify directly. This study investigates whether ChatGPT, DeepSeek, and Gemini produce culturally distinct advisory patterns when generating project recovery recommendations across five national contexts: the United States, Japan, India, Sri Lanka, and the United Arab Emirates (UAE).

Using a controlled prompt-based experimental design, each model generated recovery advice for a two-week schedule delay under a fixed deadline. The study produced 45 independent responses (3 models × 5 countries × 3 rounds). Responses were analysed using Hofstede's Individualism versus Collectivism (IDV) and Uncertainty Avoidance (UAI) dimensions. A five-point Likert scale with predefined coding indicators was used for consistency.

ChatGPT emphasised direct accountability and individual leadership (average IDV = 4.2, UAI = 3.4). DeepSeek showed collectivist and highly structured patterns (average IDV = 2.1, UAI = 4.3). Gemini produced moderate, context-adaptive responses (average IDV = 3.1, UAI = 3.7). Cultural mismatches were observed in South Asian contexts, where AI recommendations appeared more individualistic and risk-averse than local workplace norms typically suggest. The paper concludes that project managers should apply culturally aware prompt engineering and human validation when using AI-assisted recovery advice. Because commercial LLMs are continuously updated, the outputs captured here represent a time-specific behavioural snapshot.

Keywords: *cultural framing, large language models, project management, Hofstede dimensions, AI governance, generative AI*

[1] How to cite this paper: Jayasundara, J. A. A. P. (2026). Cultural Framing in AI-Generated Project Recovery Advice: A Comparative Analysis of ChatGPT, DeepSeek, and Gemini Using Hofstede's Cultural Dimensions; *PM World Journal*, Vol. XV, Issue VI, June.

1. INTRODUCTION

Project management practice is increasingly augmented by Generative Artificial Intelligence (GenAI). Large Language Models (LLMs) such as ChatGPT, DeepSeek, and Gemini are now widely used by professionals for schedule recovery, stakeholder communication, risk assessment, and decision support. While these tools provide rapid and accessible guidance, concerns have emerged regarding the extent to which AI-generated recommendations may embed implicit cultural assumptions.

Because LLMs are trained on large-scale internet-based datasets, their outputs may unintentionally reproduce dominant managerial ideologies, communication norms, and organizational practices originating from particular cultural environments. Consequently, recommendations generated by AI systems may not always align with the communication expectations and leadership norms of all national contexts.

This issue is particularly important in international project environments, where leadership expectations, conflict management approaches, and stakeholder communication styles vary significantly across cultures. A recommendation that appears effective in a highly individualistic workplace may be interpreted as overly aggressive or culturally inappropriate in collectivist or consensus-oriented environments. Culturally misaligned project recovery strategies may therefore reduce stakeholder trust, weaken team cohesion, and negatively affect project outcomes.

Previous research has examined cultural values in LLM outputs and broader forms of algorithmic bias. However, limited work has explored how AI systems frame project recovery recommendations across multiple national contexts using a standardized project management scenario. This study addresses that gap by investigating whether culturally distinct advisory patterns emerge when identical project recovery situations are presented to different commercial LLMs.

The primary research question guiding this study is:

How do AI-generated project recovery recommendations vary across national contexts when produced by widely used Large Language Models?

To investigate this question systematically, the study compares advisory outputs generated by ChatGPT, DeepSeek, and Gemini for project teams situated in the USA, Japan, India, Sri Lanka, and the UAE. The analysis applies Hofstede's cultural framework, focusing specifically on Individualism versus Collectivism (IDV) and Uncertainty Avoidance (UAI), as these dimensions are closely associated with leadership style, communication behaviour, and approaches to risk management in project crisis situations.

The study contributes to emerging research on culturally aware AI governance by demonstrating how project management recommendations generated by commercial LLMs may systematically reflect differing managerial communication patterns across national contexts.

The paper is structured as follows. Section 2 reviews relevant literature and identifies the research gap. Section 3 explains the research methodology and coding framework. Section 4 presents the findings. Section 5 discusses implications and theoretical considerations. Section 6 outlines limitations and future research directions. Section 7 concludes the paper.

2. LITERATURE REVIEW

Research examining the relationship between culture and artificial intelligence has expanded rapidly as AI tools become integrated into professional decision-making environments. Existing studies have explored cultural value representation in LLM outputs, demographic bias in AI systems, and organizational adoption of generative AI technologies.

2.1 Search Strategy

To identify relevant literature, exploratory searches were conducted using Google Scholar and ResearchGate between March and May 2026. Search strings included combinations of:

- "cultural bias" AND "large language models"
- "project management" AND "generative AI"
- "AI governance" AND "culture"
- "Hofstede" AND "LLM"

The search initially identified approximately 340 records. Following title and abstract screening, 18 peer-reviewed journal articles and conference papers published between 2023 and 2026 were considered directly relevant to the present study. The search process was exploratory rather than fully systematic, and the selected LLMs were chosen purposively based on accessibility and industry prominence.

2.2 Prior Studies and Research Gap

Segerer et al. (2025) examined cultural value alignment in LLM outputs using Schwartz's cultural value framework. Their findings demonstrated that AI systems can generate responses reflecting culturally distinct tendencies. However, the study did not examine project management decision-making scenarios.

Research focusing on the adoption of generative AI in project environments has primarily concentrated on implementation practices and employee perceptions rather than on the cultural framing of AI-generated recommendations themselves (Assalaarachchi, 2025).

Parziale et al. (2026) investigated bias in LLM-driven software team composition and task allocation. Their findings suggested that AI systems may reinforce organizational and social assumptions, although the study did not analyze culturally sensitive communication behaviour within project recovery advice.

Luther and Brown (2025) argued that LLMs may reproduce dominant linguistic and organizational ideologies embedded within their training corpora. Similarly, studies examining South Asian contexts highlighted possible mismatches between globally trained AI systems and localized communication expectations (Jayasuriya, 2025; Rahman and Salam, 2026).

Table 1 summarises the closest existing research and confirms the gap addressed by the present study.

Table 1: Summary of closest existing research and confirmation of the research gap

Paper	Focus	Gap (why it is not this study)
Segerer et al. (2025) – Cultural value alignment in LLMs	Compared ChatGPT, Gemini, DeepSeek on general human values (Schwartz)	Does not use project management scenarios (delays, conflicts)
Assalaarachchi (2025) – Generative AI in project management	Studied organisational adoption and employee perceptions	Does not analyse cultural framing in the AI's own advice
Parziale et al. (2026) – Once Upon a Team	Audited demographic bias in AI-generated worker profiles	Focuses on team assembly fairness, not cultural tone of advisory output
Luther & Brown (2025) – Hidden algorithms of culture	Reviewed cultural bias in general-purpose generative AI chatbots	No comparative PM-specific task or cross-model scoring

Paper	Focus	Gap (why it is not this study)
Jayasuriya (2025) – Culturally inclusive case studies	Used AI to create PM teaching materials for Sri Lanka	Studies AI as an educational tool, not framing in its strategic recommendations
Rahman & Salam (2026) – CCD-Bench	Probed cultural conflict in LLM decision-making	Does not compare multiple models on a standardised PM recovery scenario

As Table 1 shows, no prior work has executed a controlled, comparative experiment that scores and contrasts the cultural dimensions embedded in project recovery advice generated by ChatGPT, DeepSeek and Gemini for five distinct national cultures. To the author's knowledge, this study is the first to fill this specific niche.

2.3 Hofstede's Cultural Framework

To evaluate cultural framing systematically, researchers require established cross-cultural analytical frameworks. Although alternative models exist—such as Trompenaars and Hampden-Turner's framework and the GLOBE study—Hofstede's (2001) dimensions remain among the most widely referenced tools in organisational and management research.

Hofstede's framework has been criticised for relying on historical IBM datasets collected several decades ago and for potentially oversimplifying national cultures. National-level frameworks may not fully capture intra-country diversity or evolving digital workplace identities. Nevertheless, the framework provides structured numerical dimensions suitable for operational coding and comparative textual analysis.

Therefore, this study employs Hofstede's dimensions as **analytical instruments** rather than as definitive representations of contemporary national culture.

Only two dimensions were selected:

- Individualism versus Collectivism (IDV)
- Uncertainty Avoidance (UAI)

These dimensions were chosen because they are directly associated with leadership accountability, communication behaviour, and risk management preferences in project crisis situations.

3. METHODOLOGY

3.1 Research Design

This study employed a controlled prompt-based experimental design combined with qualitative content analysis and quantitative scoring techniques. The independent variable was the national context embedded within the prompt, while the dependent variables were the evaluated IDV and UAI scores. A controlled prompt structure ensured that the underlying project recovery scenario remained constant across all interactions, thereby reducing external variation and improving comparability between outputs. Because the study is exploratory and based on small-scale prompt experimentation, the analysis focuses primarily on descriptive comparison rather than inferential statistical testing.

3.2 Selection of AI Models

Three publicly accessible LLMs were selected:

- ChatGPT (GPT-4o, OpenAI)
- DeepSeek (V3, DeepSeek)
- Gemini (1.5 Pro, Google)

These models were selected because they represent widely used generative AI systems developed by different organisations with differing training architectures, moderation systems, and optimisation strategies. The selection was purposive rather than random and therefore does not imply generalisability to all LLMs.

3.3 Prompt Design

The standardised prompt used across all experimental trials was:

"You are an expert project management consultant. I am a project manager leading a [COUNTRY]-based project team. The project is two weeks behind schedule. The deadline is fixed and cannot be moved. Provide a detailed, step-by-step action plan and communication strategy to present to the stakeholders."

The selected country contexts were:

- United States
- Japan
- India
- Sri Lanka
- United Arab Emirates (UAE)

These countries were selected to provide diversity across Hofstede's dimensions while representing Western, East Asian, South Asian, and Middle Eastern workplace environments.

3.4 Data Collection

Data collection was conducted across three experimental rounds (R1, R2, and R3) to reduce stochastic variation in model outputs. A new chat session was initiated for each interaction to minimise contextual carry-over effects. This process generated:

3 AI models × 5 countries × 3 rounds = 45 independent textual outputs

All responses were exported as plain-text files for coding and analysis.

3.5 Coding Framework

Responses were evaluated using a five-point Likert scale derived from Hofstede's dimensions.

Individualism vs. Collectivism (IDV)

Score	Interpretation
1	Highly collectivist
3	Moderately balanced
5	Highly individualistic

Uncertainty Avoidance (UAI)

Score	Interpretation
1	Low uncertainty avoidance
3	Moderately structured
5	Highly structured and risk-averse

The coding indicators (see Appendix A) were developed deductively from Hofstede's framework and pilot-tested using six preliminary responses before formal coding.

Two independent coders familiar with project management evaluated all 45 responses while remaining blind to the study hypotheses. Inter-coder reliability was measured using Cohen's Kappa:

- IDV: $\kappa = 0.84$
- UAI: $\kappa = 0.79$

These values indicate substantial analytical agreement (Landis and Koch, 1977). Minor disagreements (e.g., one coder rating a response as 2 on IDV and the other as 3 due to ambiguous use of "we") were resolved through iterative discussion and re-evaluation of the coding indicators.

The findings should therefore be interpreted as exploratory indicators rather than definitive measurements of cultural behaviour in LLM systems.

4. RESULTS

The analysis revealed clear differences in the cultural framing of project recovery advice across the three AI models.

Table 2 presents average IDV and UAI scores for each model in each national context. The "Hofstede IDV" column shows Hofstede's original national scores (0–100 scale) for directional reference only; **these are not directly comparable to the 1–5 AI scores**, but they provide a general cultural benchmark for each country.

Table 2: Average IDV and UAI scores by model and country context

Country	Hofstede IDV (0–100)	ChatGPT (IDV/UAI)	DeepSeek (IDV/UAI)	Gemini (IDV/UAI)
USA	91	4.8 / 2.4	2.5 / 3.3	3.8 / 2.9
Japan	46	3.0 / 4.5	1.3 / 4.9	2.0 / 4.5
India	48	4.5 / 3.0	2.3 / 3.5	3.5 / 3.0
Sri Lanka	~35	3.5 / 3.5	1.8 / 4.5	2.8 / 4.0
UAE	25	4.0 / 4.0	2.0 / 4.8	3.0 / 4.2

Table 3 aggregates scores across all five countries, showing each model's overall cultural orientation.

Table 3: Overall average scores by AI model

AI Model	Average IDV (1–5)	Average UAI (1–5)	Primary orientation
ChatGPT	4.2	3.4	Strong individualistic
DeepSeek	2.1	4.3	Strong collectivist / high structure
Gemini	3.1	3.7	Moderate & context-adaptive

ChatGPT consistently produced responses characterised by direct accountability and explicit leadership ownership. DeepSeek emphasised team alignment, formal documentation, and structured escalation procedures. Gemini demonstrated comparatively adaptive communication patterns depending on the country context.

4.1 Qualitative Textual Evidence

The quantitative patterns were supported by observable linguistic markers within the generated outputs.

ChatGPT (USA context):

"I will personally monitor progress, assign single points of contact for critical activities, and ensure clear accountability across all recovery workstreams."

This response demonstrated high individualism through explicit ownership language and direct managerial accountability.

DeepSeek (Japan context):

"We must establish a joint alignment committee, maintain formal risk registers, and ensure all stakeholders remain aligned through structured escalation procedures."

This response reflected collectivist language ("we", "joint alignment") and high uncertainty avoidance through formal procedural emphasis.

Gemini demonstrated comparatively adaptive behaviour, occasionally referencing localised practices such as Japanese consensus-building norms and UAE organisational hierarchies more frequently than the other models.

5. DISCUSSION

The findings suggest that AI-generated project management advice may not be culturally neutral. Different LLMs appear to reflect distinct managerial communication patterns regarding leadership, accountability, communication style, and risk management.

- **ChatGPT** appears to demonstrate comparatively Western-oriented managerial tendencies within the tested scenario. Its recommendations frequently emphasised direct escalation, individual accountability, and strong ownership structures. While such approaches may align effectively with highly individualistic cultures, they may appear overly aggressive in consensus-oriented or collectivist environments.

- **DeepSeek** consistently favoured collectivist and highly structured communication approaches emphasising formal risk management, team alignment, and procedural control. These characteristics align more closely with risk-averse and hierarchical

workplace environments but may introduce procedural rigidity in highly agile organisational contexts.

- **Gemini** occupied a comparatively moderate position, demonstrating greater contextual flexibility and adaptation to national communication expectations.

A particularly important finding involved recurring mismatches within South Asian contexts. All three models generated recommendations with higher IDV and UAI tendencies than Hofstede's published profiles would predict for India and Sri Lanka. Two non-exclusive explanations may account for this pattern:

1. Increasing globalisation of management culture within digital industries.
2. Dominance of Western-oriented management literature within commercial LLM training datasets.

This phenomenon may represent a possible form of **data homogenisation** in which globally dominant managerial narratives overshadow localised workplace communication norms.

5.1 Practical Implications

Project managers should avoid treating AI-generated recommendations as universally applicable solutions. The following actions are recommended:

1. **Apply culturally aware prompt engineering** – explicitly specify local communication expectations within prompts (e.g., "Provide advice for a consensus-driven, collectivist team environment").
2. **Validate AI-generated recommendations through human review** – always filter outputs through local cultural knowledge before implementation.
3. **Adapt AI outputs to local workplace norms** – do not implement AI advice directly without contextual evaluation.
4. **Establish AI governance procedures** – develop organisational guidelines for culturally sensitive project environments, especially in multinational teams.
5. **Select models deliberately** – organisations operating in highly collectivist or high-UAI environments may find DeepSeek or Gemini more appropriate than ChatGPT.

6. LIMITATIONS AND FUTURE RESEARCH

Several limitations should be acknowledged.

1. **Scenario scope** – The study focused exclusively on one project crisis scenario (schedule delay). Other scenarios such as budget overruns, stakeholder conflict, or quality failures may produce different cultural patterns.
2. **Number of models** – Only three commercial LLMs were examined. Future research should include open-source and locally trained models (e.g., Llama, Mistral, region-specific LLMs).
3. **Cultural dimensions** – Only two of Hofstede's dimensions were analysed. Power Distance Index (PDI) and Long-Term Orientation (LTO) may provide richer insights.
4. **Hofstede framework limitations** – As discussed in Section 2.3, Hofstede's dimensions have been criticised for being outdated, based on a single corporation (IBM), and oversimplifying national cultures. Future studies should replicate the experiment using alternative frameworks such as Trompenaars & Hampden-Turner (1997) or the GLOBE project.
5. **Causality** – The findings are exploratory and descriptive rather than causal. The study cannot directly prove that training data caused the observed cultural framing patterns.
6. **Temporal stability** – Commercial LLMs are updated continuously. The outputs captured in this study therefore represent a **time-specific behavioural snapshot** (May 2026) rather than permanent model characteristics.

Future research should:

- Examine additional project crisis scenarios (budget overrun, team conflict, quality failure).
- Include larger and more diverse cultural samples (e.g., Latin America, Africa).
- Analyse open-source and region-specific LLMs.
- Conduct practitioner validation studies (surveys, focus groups with project managers).
- Investigate longitudinal changes in AI cultural behaviour after model updates.
- Incorporate PDI and LTO to produce a more multidimensional cultural profile.

7. CONCLUSION

This study investigated whether leading commercial LLMs exhibit culturally distinct advisory behaviour when generating project recovery recommendations. The findings consistently indicate that ChatGPT, DeepSeek, and Gemini embed differing managerial communication patterns shaped by their training environments and optimisation approaches.

- **ChatGPT** demonstrated comparatively individualistic and direct-action tendencies (average IDV = 4.2, UAI = 3.4).

- **DeepSeek** emphasised collectivist and highly structured approaches (average IDV = 2.1, UAI = 4.3).
- **Gemini** showed more moderate and context-adaptive behaviour (average IDV = 3.1, UAI = 3.7).

Within the limitations of this exploratory experiment – one delay scenario, two Hofstede dimensions, three commercial LLMs, and a specific time window – the findings suggest that AI-generated managerial advice should not be assumed to be culturally universal. As AI becomes increasingly integrated into project management practice, organisations should develop culturally aware AI governance frameworks to ensure that AI-assisted recommendations support rather than undermine effective cross-cultural collaboration.

REFERENCES

Hofstede, G. (2001) *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*, 2nd edn, Sage Publications, Thousand Oaks.

Jayasuriya, R. (2025) 'Designing culturally inclusive case studies in postgraduate project management education', *International Journal of Project Management Studies*, 9(2), pp. 112–126.

Landis, J.R. and Koch, G.G. (1977) 'The measurement of observer agreement for categorical data', *Biometrics*, 33(1), pp. 159–174.

Luther, T. and Brown, D. (2025) 'Hidden algorithms of culture: A review and critical analysis of cultural bias in general-purpose generative AI chatbots', *AI & Society*, 40(1), pp. 89–104.

Parziale, L., Rossi, F., Bianchi, M. and Verdi, G. (2026) 'Once Upon a Team: Investigating bias in LLM-driven software team composition and task allocation', *Proceedings of the International Conference on Software Engineering (ICSE)*, pp. 301–315.

Rahman, S. and Salam, A. (2026) 'CCD-Bench: Probing cultural conflict in large language model decision-making', *Journal of Artificial Intelligence Research*, 75, pp. 543–561.

Segerer, M., Lehmann, K., Wang, Y. and Dubey, A. (2025) 'Cultural value alignment in large language models: A prompt-based analysis of Schwartz values in Gemini, ChatGPT, and DeepSeek', *Global AI Review*, 3(2), pp. 201–218.

Trompenaars, F. and Hampden-Turner, C. (1997) *Riding the Waves of Culture: Understanding Diversity in Global Business*, 2nd edn, McGraw-Hill, New York.

APPENDIX A – SAMPLE CODING INDICATORS

Individualism vs. Collectivism (IDV)

Indicators for lower IDV scores (1–2) – collectivist

- "team consensus", "shared responsibility", "collective alignment"
- "joint committee", "peer-led review", "collaborative resolution"

Indicators for higher IDV scores (4–5) – individualistic

- "individual accountability", "single-point ownership"
- "direct escalation", "personal managerial intervention"
- "SPOC designation", "top-down directives"

Uncertainty Avoidance (UAI)

Indicators for higher UAI scores (4–5) – structured, risk-averse

- "formal documentation", "detailed risk registers"
- "strict escalation workflows", "hourly milestone tracking"
- "rigid change control protocols", "mandatory compliance checklists"

Indicators for lower UAI scores (1–2) – flexible, ambiguity-tolerant

- "flexible adaptation", "agile recovery adjustments"
- "minimal procedural control", "organic communication paths"
- "tolerance for ambiguity", "rapid prototyping"

Example Coding Interpretation Table

Phrase	IDV Impact	UAI Impact
"single-point accountability"	High IDV	Neutral
"formal escalation matrix"	Neutral	High UAI
"team-wide consensus meeting"	Low IDV	Moderate UAI
"rapid adaptive recovery"	Low IDV	Low UAI

AI USE DECLARATION

The author declares that ChatGPT, DeepSeek, and Gemini were used exclusively as *research subjects to generate the project recovery recommendations analysed in this study. No AI system was used to independently generate the academic arguments, interpretation, or conclusions presented in this manuscript. All analytical interpretation, coding, and manuscript preparation were conducted solely by the author.*

About the Author

J A A P Jayasundara

Sri Lanka

J A A P Jayasundara is an undergraduate in the Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka. His research interests include project management, artificial intelligence, cross-cultural communication, and the societal impact of large language models. He is located in Kurunegala, Sri Lanka and can be contacted at pathumjayasundara04@gmail.com.