Intelligent Human Systems Integration (IHSI), Vol. 200, 2026, 724–734
https://doi.org/10.54941/ahfe1007131

Intercultural Human-Centred AI for Automotive Systems: Bridging ASPICE Processes and Intelligent Human Systems Integration

Rüdiger Heimgärtner

Intercultural User Interface Consulting, Undorf, Bavaria, Germany

ABSTRACT

The integration of Artificial Intelligence into safety-critical automotive systems demands approaches that combine technical rigor with cultural and human-centred sensitivity. While Automotive SPICE (ASPICE) ensures structured engineering practices, it lacks explicit guidance on integrating intercultural human factors and intelligent AI-driven support. This paper proposes a framework for Intercultural Human-Centred AI that bridges ASPICE-based practices with intelligent human systems integration, ensuring both compliance and adaptability in global engineering contexts. Three core challenges are addressed. First, AI-driven assessments can produce inconsistent outputs, which can be addressed through structured, cache-augmented generation pipelines combined with explainable decision layers, demonstrating how consistency and auditability can be achieved in large-scale assessments. Second, engineering processes are embedded in diverse cultural contexts, requiring design principles for intercultural user interface design that support inclusive AI interaction across geographically distributed teams. Third, intelligent assistant systems are positioned as partners rather than replacements for assessors through deterministic workflow architectures that enhance decision-making while maintaining human responsibility. This paper combines computational modelling with real-world deployment evidence. Developed by an experienced ASPICE assessor, the system demonstrates how process assessments can be augmented with AI-based retrieval and reasoning. A prototype was deployed to 12 domain experts who processed 424 queries over five days, generating 218,123 tokens with high perceived usefulness ratings and strong adoption intent. This work provides both methodological foundation and practical tools for organizations integrating people and intelligent systems effectively, with relevance beyond automotive to all regulated industries requiring compliance, cultural inclusivity, and human-AI collaboration.

Keywords: Human-centred AI, Automotive SPICE, Automotive systems, Intercultural design, Cache-augmented generation, Process assessment, ASPICE advisor system

INTRODUCTION

Motivation and Context

The automotive industry is undergoing fundamental transformation driven by software-intensive systems, artificial intelligence, and global collaboration. Modern vehicles have evolved from purely mechanical products into

Received September 20, 2025; Revised November 10, 2025; Accepted November 25, 2025; Available online February 1, 2026

complex cyber-physical systems where software and AI-driven functionalities determine safety, user experience, and competitive differentiation. This transformation places unprecedented demands on engineering processes, particularly in ensuring quality, safety, and regulatory compliance across internationally distributed development teams.

Automotive SPICE (ASPICE), based on ISO/IEC 33020, has emerged as the de facto standard for process assessments in automotive software development. Its Process Assessment Model (PAM) Version 4.0 provides a comprehensive framework covering 43 processes across acquisition, supply chain, engineering, support, management, improvement, and reuse domains.

Process assessments require consistent interpretation of capability indicators across projects, assessors, and organizational contexts. Human assessors, while experts (or precisely for that reason), introduce variability through cognitive biases, varying interpretations, and differing levels of domain expertise. I.e., interpretation objectivity seems to be not possible. AI-based assistance can be a means to reduce this variability by increasing the interpretation objectivity, but only if the AI systems themselves are consistent, explainable, and auditable—requirements particularly stringent in safety-critical automotive contexts governed by ISO 26262 and related standards.

Automotive development increasingly involves geographically distributed teams spanning diverse cultural contexts, from German OEMs to Chinese suppliers, from Indian software houses to American engineering centres. Cultural differences influence not only communication styles but also fundamental assumptions about process rigor, documentation practices, and quality expectations (Hoffmann et al., 2004). ASPICE provides process requirements but lacks guidance on adapting assessment and improvement practices to intercultural contexts.

As AI systems become more capable, a critical question emerges regarding how human expertise and AI capabilities should be combined in process-critical activities like ASPICE assessments. The automotive sector requires approaches that leverage AI efficiency while maintaining human oversight, responsibility, and domain judgment, particularly in decisions affecting safety certification and regulatory compliance.

Research Gap and Contribution

While significant research exists in adjacent domains such as AI explainability (Arrieta et al., 2020; Saeed & Omlin, 2023), human-centred AI design (Shneiderman, 2020; Haque et al., 2023), intercultural human-centred interaction (HCI) (Heimgärtner 2012) and intercultural user interface design (IUID) (Heimgärtner, 2019), and AI-assisted software engineering (Nascimento et al., 2023), the intersection of these domains—particularly in the context of process assessment automation—remains underexplored (Edwar et al., 2022).

This paper addresses this gap through proposing an integrated framework that bridges ASPICE process assessment with cache-augmented generation techniques, integrates intercultural design principles into AI-assisted engineering workflows, demonstrates a deterministic workflow architecture

allowing reproducibility and auditability regarding functional equivalence, and validates the approach through real-world deployment with domain experts.

RELATED WORK

Process Assessments and ASPICE

Automotive SPICE originates from the Software Process Improvement and Capability Determination framework defined in ISO/IEC 15504, now superseded by ISO/IEC 33000 series standards. The current Automotive SPICE PAM Version 4.0 (released by the VDA Quality Management Centre) defines 43 processes with capability levels ranging from 0 (incomplete) to 5 (innovating). Traditional ASPICE assessments are conducted by certified assessors who evaluate projects against base practices, work products, and generic practices. Research on ASPICE has primarily focused on organizational maturity models, assessment method comparisons, and integration with agile methodologies, but limited work addresses AI-assisted assessment support. The potential of large language models (LLM) for software engineering process support has been explored in code generation and requirements engineering (Chen et al., 2021; Jiang et al., 2023), but not specifically for process capability assessments.

Human-Centred AI and Explainability

Human-Centred AI emphasizes designing AI systems that augment rather than replace human capabilities (Shneiderman, 2020). Key principles include maintaining human control, ensuring transparency, and designing for diverse users. In safety-critical domains, explainability becomes paramount not merely to satisfy user curiosity but to enable compliance verification and error diagnosis. Recent work on retrieval-augmented generation by Lewis et al. (2020) addresses some explainability concerns by grounding generation in retrievable sources, though it introduces latency and retrieval errors. Emerging research on cache-augmented generation (Chan et al., 2024) demonstrates that preloading knowledge into extended context windows can eliminate retrieval latency while maintaining accuracy, particularly relevant for bounded domains like ASPICE with manageable knowledge bases.

Intercultural Design and Human-Autonomy Teaming

Intercultural user interface design addresses how cultural differences influence HCI design (Heimgärtner, 2019). Hofstede's cultural dimensions provide one framework for understanding these differences, though prior work has examined cultural adaptation primarily in consumer applications rather than engineering tool design. The author's work (Heimgärtner, 2019) established design principles for intercultural interfaces in technical domains, demonstrating that cultural factors significantly influence acceptance and effectiveness of engineering support tools.

Human-autonomy teaming research investigates how humans and AI systems can collaborate effectively, with key challenges including appropriate

reliance, transparency of AI capabilities, and maintaining human situation awareness. Multi-agent architectures have gained attention for complex problem-solving, but recent research highlights failure modes including communication overhead, coordination challenges, and non-deterministic behaviour. For compliance-critical applications like ASPICE assessment, determinism and traceability are essential (Heimgärtner, 2025; Huang et al., 2024).

This work synthesizes these domains by applying cache-augmented generation techniques specifically to ASPICE process assessment, integrating intercultural user interface design principles into AI assistant design for globally distributed engineering teams, and demonstrating deterministic workflow architectures outperform multi-agent approaches in compliance-sensitive contexts.

FRAMEWORK FOR INTERCULTURAL HUMAN-CENTERED AI

Three Glue-Dimensions

The framework addresses the integration of AI into ASPICE-based automotive engineering through three interdependent dimensions that must be considered simultaneously for successful deployment (Figure 1).

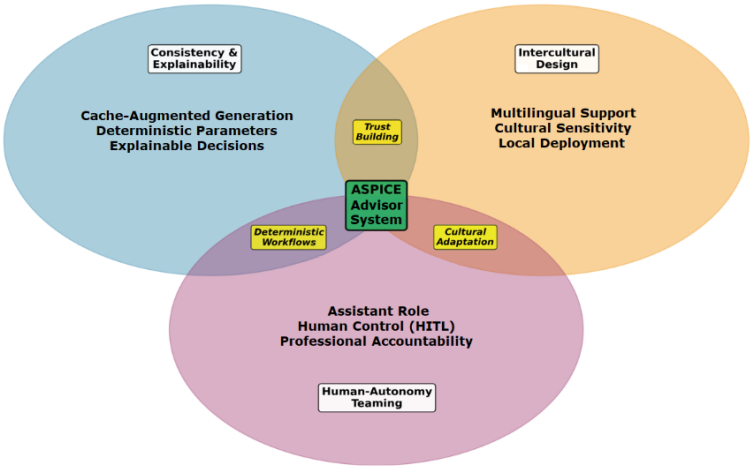

Figure 1: Preliminary three-dimensional framework for intercultural human-centred AI. The framework integrates consistency & explainability, intercultural design, and human-autonomy teaming with the ASPICE advisor system at their intersection.

The "consistency and explainability dimension" ensures that AI-assisted assessments produce reliable, auditable results across different contexts and assessors. The "intercultural design dimension" addresses the reality that engineering teams operate in diverse cultural contexts with different expectations, communication styles, and trust requirements. The "human-autonomy teaming dimension" defines how human experts and AI capabilities should be combined to leverage the strengths of both while maintaining appropriate human oversight and responsibility.

SYSTEM ARCHITECTURE

Technical Foundation

The ASPICE Advisor System developed by the author implements the framework through a local deployment architecture running on commodity hardware. The core component is an open-source Large Language Model with 128k token context window, chosen for balance of capability, efficiency, and extended context support. Local deployment ensures data sovereignty, enables offline operation, and provides full control over generation parameters and knowledge base content while the system achieves enterprise-grade reliability. Furthermore, this democratizes access to advanced AI capabilities also for organizations with security or resource constraints.

The knowledge base comprises the complete ASPICE PAM 4.0 documentation chunked into semantically coherent segments with overlapping to preserve context across boundaries. The chunks are embedded using a vector embedding model with enough dimensions to cope with the chunk size, providing an optimal balance between computational efficiency and semantic accuracy for bounded knowledge domains.

Deterministic Workflow Pipeline

The generation pipeline implements enough determinism through multiple mechanisms to ensure deterministic functional equivalence workflows in contrast to multi-agent-systems. A fixed random seed sets the ground for reproducible pseudo-random generation across sessions with an LLM. Temperature is set to zero, forcing the model to always select the highest-probability token rather than sampling from a probability distribution. These parameters represent the primary basis for the desired result that given identical inputs, the system produces identical outputs at least in terms of functional equivalence, enabling reproducibility essential for audit trails. The six-step process ensures reproducible outputs through fixed random seed, zero temperature, and greedy decoding, with GPU acceleration reducing embedding time from 500ms to 50ms per query.

IMPLEMENTATION AND DEPLOYMENT

Real-World Beta Deployment

A five-day beta deployment of the ASPICE Advisor System with Chatbot-GUI from October 23–28, 2025, provided real-world validation with 12 verified domain experts including ASPICE assessors, automotive engineers, and process improvement specialists. The system operated 100% on time and zero error rate across 424 queries demonstrates possible production-level reliability. Total token processing reached 218,123 tokens with an average of 514 tokens per query, indicating substantive interactions rather than trivial testing. The users could ask the system for real-time support during assessments. The most frequently applied use case was to get answers regarding the content of the ASPICE PAM 4.0 (e.g. what are the base practices for process X or what are reasonable metrics for it).

User Feedback and Validation

An integrated survey collected 46 qualitative responses providing insights into perceived usefulness and adoption intent. Usefulness ratings clustered in the 7–9 out of 10 range (mean: 8.67), indicating strong value perception among domain experts. Representative feedback highlighted specific benefits including rapid access to ASPICE process details, consistent interpretation guidance, and time savings compared to manual documentation searches.

DISCUSSION

Interpretation of Results

The beta deployment results validate the suggested three-dimensional framework for Intercultural Human-Centered AI in ASPICE contexts above. The high usefulness ratings (mean: 8.67/10) and strong adoption intent (58% definite, 42% considering) demonstrate that domain experts perceive significant value in AI-assisted process assessments. The sustained engagement pattern with 424 queries over five days and substantial token generation (218,123 tokens) indicates that users found the system sufficiently useful to integrate it into their actual work processes rather than treating it as curiosity.

The system's performance metrics reveal important insights into the technical approach. The 100% uptime and zero error rate during the deployment period demonstrate the reliability benefits of the deterministic workflow architecture. This contrasts with typical multi-agent systems where coordination failures and non-deterministic behavior can introduce unpredictable failure modes. For safety-critical automotive contexts where audit trails and reproducibility are essential, this quasi-deterministic behavior represents a crucial advantage over more complex architectures.

The cache-augmented generation approach proved particularly well-suited for ASPICE assessment support. By preloading the complete PAM 4.0 documentation into the extended context window, the system eliminated retrieval latency while maintaining comprehensive coverage of all 43 processes. The bounded nature of standards such as ASPICE PAM 4.0—while extensive, it is finite and well-defined—made it an ideal candidate for this approach. This challenges the default assumption that retrieval-augmented generation is always necessary for knowledge-intensive tasks, particularly when the knowledge base is stable and manageable in size. However, if it comes to applying the ASPICE knowledge (not only looking up pieces of information, but rating or consulting), much experience is necessary and must be covered by extending the knowledge base extensively.

Implications for Practice

For organizations conducting ASPICE assessments, the results suggest several practical implications. First, AI assistance can meaningfully augment—not replace—assessor capabilities when designed with appropriate human-autonomy teaming principles. The system's role as a rapid reference and

consistency check supports assessors in their decision-making without attempting to automate the judgment-intensive aspects of capability determination.

Second, local deployment addresses critical data sovereignty concerns in automotive contexts where assessment information may be confidential or proprietary. The ability to operate entirely offline while maintaining full functionality enables deployment in security-conscious environments without compromising on AI capabilities. This architectural decision reflects real-world organizational requirements rather than purely technical optimization.

Third, the intercultural design considerations proved essential for acceptance across the international participant base. Support for multilingual interaction, culturally appropriate communication styles, and professional terminology alignment helps build trust and reduce friction in adoption. Organizations deploying AI assistance in global engineering contexts should consider these factors as first-class requirements rather than afterthoughts.

Comparison With Related Work

Compared to multi-agent architectures for complex problem-solving, the deterministic workflow architecture regarding functional equivalence prioritizes reliability and auditability over architectural flexibility. While multi-agent systems can handle more open-ended tasks, recent research (Huang et al., 2024) has highlighted their failure modes in production deployments such as communication overhead, coordination challenges, and non-deterministic behavior. For compliance-critical applications like ASPICE assessment, these risks outweigh potential benefits, making deterministic workflows the more appropriate choice.

The intercultural design aspects of this work extend existing human-centered AI principles (Shneiderman, 2020) into the specific context of international engineering collaboration. While previous work on intercultural HCI (Heimgärtner, 2019) established design principles for technical domains, this deployment demonstrates their practical application in AI-assisted assessment tools. The positive user feedback across diverse cultural backgrounds validates the importance of these considerations.

Broader Significance

Beyond ASPICE and automotive contexts, this work contributes to a broader understanding of how AI should be integrated into professional knowledge work. The three-dimensional framework provides a template applicable to other regulated industries where compliance, cultural diversity, and human expertise must coexist with AI assistance.

LIMITATIONS

Deployment Scope and Duration

The five-day beta deployment, while providing valuable insights, represents a limited temporal window for evaluating long-term adoption patterns and

usage evolution. Participants may have exhibited novelty effects or selection bias, as early adopters often demonstrate higher engagement than the broader user population. Extended longitudinal studies would be needed to assess sustained usage patterns, workflow integration, and identification of edge cases that emerge only with prolonged use.

The participant pool of 12 domain experts, while representing diverse roles and backgrounds, constitutes a relatively small sample size for statistical generalization. In addition, they have been selected randomly by the author out of his own expert network which possibly raised a bias towards positive feedback. The concentration of usage among the top five users (92% of queries) further limits the representativeness of the findings. Larger-scale deployments across multiple organizations and assessment contexts with participations unknown to the author would strengthen the evidence base and reveal organizational or contextual factors affecting adoption and effectiveness.

Technical Constraints

The cache-augmented generation approach, while effective for ASPICE PAM 4.0's bounded knowledge base, faces scalability limitations. As context windows extend, the approach can accommodate larger knowledge bases, but practical limits remain for truly massive or rapidly evolving corpora. Organizations with knowledge bases exceedingly even extended context window capacities would need to employ hybrid approaches combining caching with selective retrieval.

The system's reliance on a single large language model introduces dependency on that model's capabilities and limitations. Architectural constraints, hallucination tendencies, and reasoning limitations inherent to current LLM technology affect the system's performance. While the deterministic workflow and structured knowledge base mitigate some risks, fundamental limitations of transformer-based architecture remain relevant. Future deployments would benefit from architectural diversity or ensemble approaches.

Local deployment on commodity hardware introduces resource constraints that may not scale to very large user populations or extremely high query loads. The system's throughput and latency characteristics under peak load conditions were not fully characterized during the deployment period. Production deployments would require capacity planning and potentially infrastructure scaling considerations.

Evaluation Methodology

The evaluation relied primarily on self-reported usefulness ratings and adoption intent rather than objective performance metrics such as assessment accuracy, time savings, or quality improvements. While perceived usefulness correlates with actual benefit, independent validation of assessment quality and consistency improvements would strengthen the evidence. Comparative studies against traditional assessment processes would provide more rigorous effectiveness measures.

The qualitative feedback collected, while valuable for understanding user perspectives, was not analyzed using systematic qualitative research methodologies. Thematic analysis, grounded theory, or other structured approaches to qualitative data could reveal deeper insights into usage patterns, pain points, and improvement opportunities. The 46 survey responses, though informative, represent only a subset of total interactions and may not capture the full range of user experiences.

Generalizability

The findings emerge from ASPICE assessment contexts in automotive engineering provides a template applicable to other regulated industries. However, generalization to other process assessment frameworks or domains require careful consideration. Empirical validation in other regulated industries, process standards, or cultural contexts remains necessary. Different process frameworks may present different knowledge structures, assessment practices, or cultural dynamics that affect system design and effectiveness.

The intercultural design principles are grounded in established cross-cultural HCI research. However, they were implemented for a specific participant demographic and may not address all cultural dimensions relevant in global engineering contexts. The deployment's geographic and organizational diversity does not represent the full spectrum of cultural variation in international automotive development. Different cultural combinations or organizational contexts might reveal additional design requirements or adoption barriers.

Knowledge Currency and Maintenance

The system's knowledge base, preloaded with ASPICE PAM 4.0 documentation, requires updating when new process model versions release or when organizational interpretations evolve. The cache-augmented approach's reliance on preloaded knowledge introduces maintenance overhead compared to dynamic retrieval systems that can adapt automatically to knowledge base updates. Organizations must establish governance processes for knowledge base updates, version control, and validation of modified content. The evaluation did not assess the system's behavior with outdated, incomplete, or contradictory knowledge. Real-world deployments may encounter scenarios where documentation versions conflict, interpretations differ across organizational units, or process tailoring introduces local variations. The system's ability to handle such complexity and maintain consistency across variant contexts requires further investigation.

CONCLUSION

This paper suggested a framework for Intercultural Human-Centered AI applied to ASPICE-based automotive product engineering. The intersection of process standardization, cultural diversity, and AI assistance through a three-dimensional framework emphasizing consistency, intercultural design, and human-autonomy teaming was addressed. The ASPICE Advisor

system demonstrated these principles through cache-augmented generation, deterministic workflows, and culturally aware interface design.

Real-world deployment with 12 domain experts processing 424 queries validated the approach, showing high perceived usefulness and adoption intent. The deterministic architecture provided audit trails essential for compliance while the local deployment addressed data sovereignty concerns. The system supported multilingual interaction and professional communication norms, building trust across cultural contexts.

Future work must address the identified limitations through extended deployments across multiple organizations, objective assessment quality metrics, and hybrid approaches combining caching with selective retrieval for larger knowledge bases. Additionally, systematic evaluation of knowledge currency maintenance processes and the system's behaviour under peak load conditions remains essential for production-grade deployments.

ACKNOWLEDGMENT

The author would like to acknowledge the twelve domain experts who participated in the five-day beta deployment of the ASPICE Advisor System with Chat-Bot-Style-User-Interface from October 23–28, 2025.

REFERENCES

Arrieta, A.B., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.

Chan, S., et al. (2024). Cache-augmented generation: Eliminating retrieval latency through context preloading. Proceedings of the ACM Web Conference 2025 (WWW '25). https://doi.org/10.1145/3701716.3715490

Chen, M., et al. (2021). Evaluating large language models trained on code. arXiv preprint arXiv:2107.03374.

Edwar, E., Sameh, S., & Sobh, I. (2022). ASPICE Applicability on New Automotive Technologies (AI). In EuroSPI 2022. Communications in Computer and Information Science, vol 1646. Springer. https://doi.org/10.1007/978-3-031-15559-8_31

Hall, E.T., & Hall, M.R. (1990). Understanding Cultural Differences: Germans, French and Americans. Intercultural Press.

Haque, A.B., et al. (2023). Explainable Artificial Intelligence (XAI) from a user perspective. Technological Forecasting and Social Change, 186, 122120.

Heimgärtner, R. (2012). Cultural Differences in Human-Computer Interaction: Towards Culturally Adaptive Human-Machine Interaction. München: Oldenbourg Wissenschaftsverlag. https://doi.org/10.1524/9783486719895

Heimgärtner, R. (2019). Intercultural User Interface Design. Springer International Publishing.

Heimgärtner, R. (2025). Human-Centered AI for Automotive Systems: Towards Explainable, Intercultural, and Standardized Integration. In: Tareq Ahram, Waldemar Karwowski and Jay Kalra (eds) Human Factors in Design, Engineering, and Computing. AHFE (2025) International Conference. AHFE Open Access, vol 199. AHFE International, USA. http://doi.org/10.54941/ahfe1006935

Heimgärtner, R., et al. (2011). Empathy as Key Factor for Successful Intercultural HCI Design. In: Marcus, A. (eds) Design, User Experience, and Usability. DUXU 2011. Lecture Notes in Computer Science, vol 6770. Springer.

Hoffmann, H., Schoper, Y., & Fitzsimons, C. (2004). Internationales Projektmanagement: Interkulturelle Zusammenarbeit in der Praxis. München: Beck im dtv.

Hofstede, G. (2010). Culture's Consequences: International Differences in Work-Related Values (2nd ed.). Sage Publications.

Huang, J.-T., et al. (2024). On the Resilience of LLM-Based Multi-Agent Collaboration with Faulty Agents. arXiv preprint arXiv:2408.00989.

ISO/IEC 33020:2015. Information technology—Process assessment—Process measurement framework for assessment of process capability.

Jiang, N., et al. (2023). Large language models for requirements engineering. IEEE Software, 40(3), 54–61.

Lewis, P., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.

Lewis, R.D. (2006). When Cultures Collide: Leading Across Cultures (3rd ed.). Nicholas Brealey International.

Nascimento, N., Alencar, P., & Cowan, D. (2023). Artificial Intelligence versus Software Engineers. In ACM Conference on Software Engineering and AI (SEAI).

Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey. Knowledge-Based Systems, 263, 110273.

Shneiderman, B. (2020). Human-centered artificial intelligence. International Journal of Human-Computer Interaction, 36(6), 495–504.

VDA QMC. (2023). Automotive SPICE Process Assessment Model v4.0. Verband der Automobilindustrie Quality Management Center.