Embedding Responsible AI Through Multidimensional Benchmarking and Organizational Learning

Bibhas Mondal[1], Parama Bhaumik[2] and Siuli Roy[3]
[1]Antarpath Impact Consulting, Kolkata, India
[2]Department of Information Technology at Jadavpur University, Kolkata, India
[3]Department of Information Technology at Heritage Institute of Technology, Kolkata, India

mebibs@gmail.com
bhaumikparama@gmail.com
siuli.roy@gmail.com

Abstract: As artificial intelligence (AI) becomes embedded in organizational decision-making, firms must balance algorithmic efficiency with human expertise and learning. While AI improves prediction and automation (Davenport & Ronanki, 2018), opaque machine learning systems may reduce interpretability and weaken professional judgment (Burrell, 2016). Research on AI in management further shows that AI reshapes rather than replaces expertise, creating tensions between automation and augmentation (Raisch & Krakowski, 2021). However, most AI governance and explainability approaches emphasize transparency principles or technical metrics without integrating them into ongoing organizational learning processes. This study proposes a multi-dimensional benchmarking framework that embeds responsible AI within organizational learning systems. The framework focuses on four dimensions: (1) explainability metrics assessing both technical transparency and user comprehension; (2) systematic bias and fairness evaluation; (3) robustness testing under uncertainty; and (4) structured human-in-the-loop validation that records expert overrides and incorporates them into retraining cycles. Unlike traditional model evaluation centred on predictive accuracy, or periodic governance audits, benchmarking is conceptualized as a continuous socio-technical feedback mechanism that supports adaptive capability and preserves human epistemic agency. Methodologically, the study combines comparative case studies of AI-enabled decision systems, simulation experiments comparing evaluation approaches, and longitudinal organizational data analysis examining override patterns and learning effects. The primary contribution is reframing benchmarking as an institutional governance architecture that connects AI performance, human oversight, and organizational learning. By embedding evaluation within knowledge processes, the framework advances prior governance and explainability models toward sustainable, human-centred AI integration.

Keywords: Responsible AI, Organisational learning, Multidimensional benchmarking, Human-in-the-loop, Explainability, Knowledge management

1. Introduction

Knowledge management (KM) is widely recognized as vital for organizational performance and competitiveness (Gelashvili-Luik et al. 2025). In the digital era, enterprise KM systems (KMS) use technology to capture and share knowledge assets (Alavi and Leidner 2001). The recent wave of AI – including machine learning and generative models – promises to *transform* KM by enabling real-time knowledge flows and more personalized knowledge discovery (Gelashvili-Luik et al. 2025). For example, Accenture's "Knowledge Assist" system (developed with AWS) uses generative AI to allow natural-language Q&A over vast corporate knowledge repositories (Geller et al. 2023). Reportedly, this has halved new-hire training time and cut support ticket escalations by approximately 40 percent (Geller et al. 2023).

Despite these successes, integrating AI into KM presents new challenges. Researchers note that AI can automate routine tasks but must be balanced with human oversight (Gelashvili-Luik et al. 2025). Key issues include data governance, algorithmic transparency, and alignment with existing processes (Gelashvili-Luik et al. 2025; Rezaei 2025). Moreover, many organizations struggle to realize AI's full potential in KM, pointing to gaps in practice and research (Gelashvili-Luik et al. 2025). This paper addresses these gaps by (1) reviewing the state of AI in KM, (2) analysing Accenture's AI-KM approach as a case study, and (3) developing a benchmarking framework to evaluate and guide AI-driven KM initiatives.

The rest of the paper is organized as follows. Section 2 reviews relevant literature on AI-enabled KM and benchmarking. Section 3 describes our case-study methodology. Section 4 details the Accenture case and identifies gaps in its AI-KM system. Section 5 presents the proposed benchmarking framework. Section 6 discusses future research directions. Section 7 concludes.

2. Literature Review

Studies consistently find that while AI can greatly enhance KM, it also introduces complex challenges. Gelashvili-Luik et al. (2025) show that successful AI-KM depends on leadership support, adaptive governance, and

technology tailored to context. They note AI's role is shifting from backend automation to enabling dynamic knowledge flows, but warn that maintaining transparency and human oversight is critical. Similarly, Nguyen et al. (2025) highlight *trust and ethics* as fundamental in AI–KM: they identify data anonymity, bias prevention, and collaborative approaches as gaps that must be bridged to make AI-driven KM transparent and inclusive. Their review emphasizes that "trust emerges as a key factor" across technological, organizational, and interpersonal dimensions (Nguyen et al. 2025).

Rezaei (2025) categorizes AI-KM challenges into technological, organisational, and ethical domains. For technology, issues include data availability, scalability, algorithm complexity, and security. Organizationally, common barriers are resistance to change, lack of expertise, and insufficient collaboration. Ethically, bias, accountability, privacy, and job displacement fears arise. In practice, each KM process phase faces specific challenges: knowledge creation may be hindered by cultural inertia and job fears; knowledge storage by inadequate skills and privacy risks; sharing by security and trust issues; and application by uncertainty and resource constraints. A subsequent industry review confirms that data security/privacy and employee concerns are among the top challenges in every phase of KM with AI.

Established KM maturity models (e.g. APQC, TSIA, KM3) stress that people, process, and technology must align (Coveo 2023). The TSIA 3.0 model, for example, explicitly incorporates generative AI and measures capabilities like knowledge base integration, AI governance, and performance metrics. Case examples of KM transformation (e.g. Tyler Technologies, athenahealth) show that firms track metrics such as case deflection rates and knowledge usage to demonstrate value. In general, KM performance is assessed both *efficiency*-wise (e.g. search time) and *effectiveness*-wise (e.g. reuse of knowledge) (Minonne and Turner 2018).

Benchmarking in KM has been advocated as a critical practice. Szydełko and Szydełko (2025) argue that benchmarking should be "an integral component of the knowledge management system" to identify gaps and implement best practices. They map out how data sources for benchmarking can be gathered across stages of KM implementation. Importantly, they highlight the *triangulation principle*: effective KM benchmarking requires using multiple data sources, expert judgments, and diverse methods in parallel. This means comparing internal KPIs with external best-practices and involving cross-functional teams, thereby increasing the reliability of any KM evaluation.

In sum, the literature illustrates AI's promise for KM but also warns of non-trivial hurdles. Key themes are: the need for robust data management, the centrality of user trust, and the socio-technical nature of AI-KM projects (Gelashvili-Luik et al. 2025; Rezaei 2025; Nguyen et al. 2025). There is also recognition that systematic measurement – via maturity models and benchmarking – is needed to guide KM strategy. Our work builds on these insights by applying them to a concrete case and deriving a dedicated benchmarking framework for AI-enabled KM.

3. Methodology

We adopt a qualitative case-study approach centred on Accenture's AI-driven KM system. Thematic analysis was conducted in three stages following Braun and Clarke (2006): (1) familiarisation with the data corpus; (2) generating initial codes from recurring concepts and challenges; and (3) collating codes into overarching themes (e.g. transparency gaps, data governance, tacit knowledge capture). These themes were then mapped against established KM frameworks to identify gaps in Accenture's approach. The secondary sources consulted for this study comprised: (a) Accenture's publicly available documentation and the AWS Machine Learning Blog post by Geller et al. (2023) describing the Knowledge Assist implementation; (b) peer-reviewed academic articles retrieved from Scopus and Web of Science using the search terms "AI knowledge management," "responsible AI benchmarking," and "organisational learning AI" (yielding, inter alia, Gelashvili-Luik et al. 2025; Nguyen et al. 2025; Rezaei 2025); (c) industry reports and KM benchmarking documentation, including Coveo (2023) on KM maturity models and TSIA Technology and Services 3.0 framework material; and (d) publicly available KM transformation case studies of Tyler Technologies and athenahealth as reported in Coveo (2023). This process follows the triangulation method suggested by Szydełko and Szydełko (2025), combining multiple information sources and expert perspectives to enhance validity. The case study provides an empirical basis for identifying gaps in AI-driven KM and for constructing our benchmarking framework.

4. Case Study: Accenture's AI-driven KM System

Accenture has been a leader in KM. A study by Paik and Cjnoi (2005) notes that by 2005, Accenture had invested over $500 million and 500 dedicated staff to its global KM platform, Knowledge Xchange. That study also

observed limitations of Accenture's early KM: its "one size fits all" global strategy did not fully account for local practices and did not adequately motivate all employees.

AI-Enabled KM: More recently, Accenture has layered AI on top of its KM infrastructure. In 2023, Accenture and AWS launched *Knowledge Assist*, a generative-AI solution for enterprise KM (Geller et al. 2023). This system ingests vast amounts of internal and external unstructured content (document databases, FAQs, etc.) and enables users to pose natural-language questions. Instead of relying on keyword search, Knowledge Assist provides concise answers by understanding context within the entire knowledge corpus (Geller et al. 2023). Notably, Accenture reports that this AI bot led to significant efficiency gains: new-employee training time was reduced by approximately 50 percent, and support escalation rates dropped by approximately 40 percent (Geller et al. 2023). The solution uses AWS's generative AI services (e.g. Claude and Titan) to process queries and continuously improves as more interactions occur.

Operational Aspects: Figure 1 (below) illustrates the system architecture described by Accenture. The key feature is "conversational ingestion": employees can ask questions in natural language, and the AI returns answers with citations. Over time, machine learning refines the answer quality. Geller et al. (2023) describe this as "transforming how knowledge is captured, organized, and used across an organization." In effect, Accenture's AI-KM extends its Knowledge Xchange by making the stored knowledge *accessible on demand*.

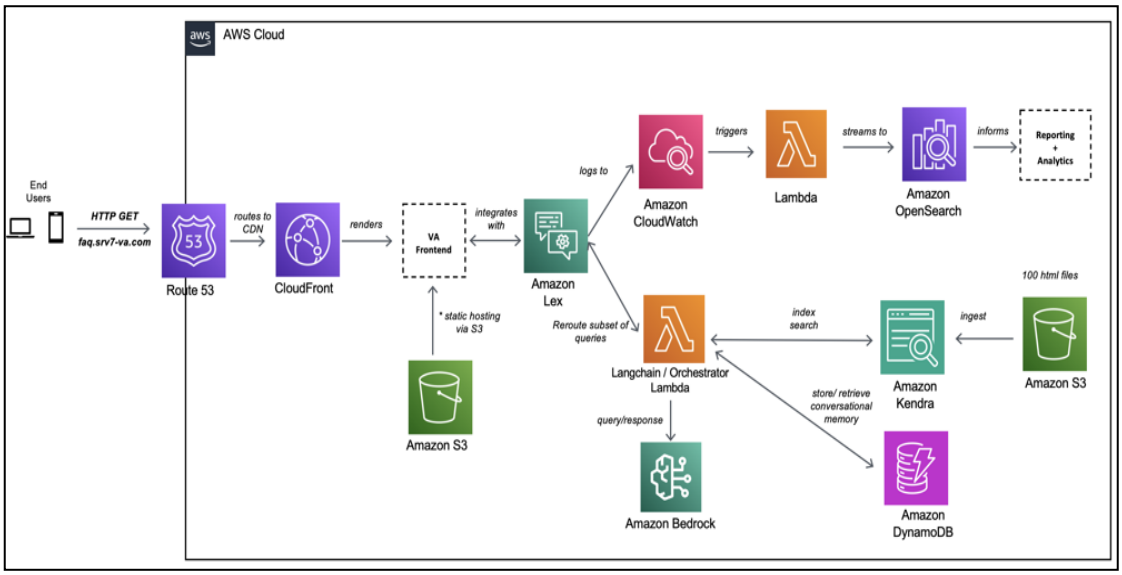


Figure 1: Accenture's AI-driven KM architecture (source: Geller et al. 2023). Key components: (1) Knowledge Assist AI engine powered by AWS generative AI services (including Claude and Titan models), which processes natural-language queries; (2) Knowledge Xchange repository, ingesting unstructured content from document databases, FAQs, and internal reports; and (3) a conversational user interface through which employees submit queries and receive cited, context-aware answers. Bidirectional feedback loops indicate that user interactions continuously refine model outputs over time

Results: According to the published case, the Knowledge Assist solution has improved knowledge access significantly (Geller et al. 2023). For example, field personnel and clients can now retrieve answers swiftly without sifting through documents. Accenture claims the AI bot provides "game-changing" benefits in connecting people to needed information (Geller et al. 2023). These results illustrate the potential of AI in enterprise KM when properly deployed. However, as we discuss next, the case also reveals certain limitations.

4.1 Identified Gaps in Accenture's AI-KM

While Accenture's AI-KM system demonstrates clear benefits, our analysis uncovered several gaps where best practices are not fully addressed. These gaps align with the challenges noted in the literature:

Data Integration and Quality: AI systems depend on high-quality data. The "Data Chaos" loop described by Gelashvili-Luik et al. (2025) warns that rapidly increasing data without proper management can degrade knowledge quality. In the Accenture case, a vast document repository is used, but it is unclear how outdated or duplicate content is pruned. Without careful curation, the AI's answers could be based on stale or inconsistent

information. This gap in data governance could erode trust and effectiveness over time (Gelashvili-Luik et al. 2025).

Transparency and Trust: Generative AI can produce opaque outputs. Neither Accenture's blog nor our sources detail mechanisms for answer explainability beyond listing sources. Gelashvili-Luik et al. (2025) emphasize that "lack of transparency" in AI can breed distrust among users. Similarly, Nguyen et al. (2025) identify trust as a *key factor* affecting all KM processes. If users find occasional AI hallucinations or errors (common in LLMs), they may lose confidence in the system. Therefore, governance frameworks with audit logs and clear accountability (as recommended by Nguyen et al. 2025) are needed but not evident in the Accenture system description.

Tacit Knowledge Capture: Accenture's solution relies on its database of explicit knowledge (documents, reports, FAQs). However, human expertise often resides as tacit knowledge (experiential insights) that is hard to codify. Gelashvili-Luik et al. (2025) note that AI tools primarily handle explicit knowledge, leaving tacit knowledge underutilized. The Accenture case does not describe any initiative to extract insights from employee experience or informal networks. This is a gap: without mechanisms (like expert communities or structured interviews) to feed tacit knowledge into the AI, valuable organizational know-how may be overlooked (Nguyen et al. 2025).

Cultural and Change Management: Introducing AI changes how people access knowledge. Rezaei (2025) warns that organizational resistance and lack of training can impede AI-KM adoption. In Accenture's global organization, ensuring that all employees adopt the new tool requires change management. The case does not detail how Accenture encouraged usage or mitigated fear of AI (for instance, fear of replacing human roles). If employees default to the old system or distrust AI answers, the system's impact will be limited. Leaders must cultivate a knowledge-sharing culture and provide training on the AI system (Paik and Cjnoi 2005), yet such efforts are not described.

Performance Metrics: Effective KM initiatives track outcomes with clear KPIs. Besides the anecdotal 50 percent and 40 percent improvements (Geller et al. 2023), Accenture has not published systematic metrics for the AI-KM. Szydełko and Szydełko (2025) emphasize benchmarking benefits (e.g. ticket deflection rates, user satisfaction) to guide improvement. Without such metrics, it is hard to gauge whether the AI assistant truly enhances knowledge reuse or simply speeds specific tasks. The lack of disclosed KPIs or A/B testing data for the KM bot is a notable gap in accountability.

Ethics and Privacy: Generative AI raises concerns about data privacy and bias. For example, the AI might inadvertently reveal sensitive information if not properly constrained. Gelashvili-Luik et al. (2025) and Nguyen et al. (2025) call for responsible AI governance in KM. The Accenture description focuses on technology and efficiency, but does not discuss safeguards (e.g. content filtering, compliance). Given that some content may be proprietary or personal (e.g. client-specific documents), appropriate anonymization or access controls are needed. The absence of detail on privacy measures is a gap that aligns with broader industry calls for ethics in AI (Nguyen et al. 2025).

In summary, while Accenture's AI-KM demonstrates advanced technical capabilities, several socio-technical aspects need strengthening. The identified gaps – in data management, trust, tacit knowledge, organizational readiness, metric tracking, and ethics – mirror the challenges reported in academic studies (Gelashvili-Luik et al. 2025; Rezaei 2025; Nguyen et al. 2025). The next section uses these insights to develop a benchmarking framework that organizations like Accenture can use to assess and improve their AI-driven KM systems.

5. Benchmarking Framework for AI-driven KM

We propose a five-dimensional framework for evaluating AI-enabled KM, informed by KM benchmarking research (e.g. Szydełko & Szydełko 2025) and prior KM studies (Gelashvili-Luik et al. 2025; Nguyen et al. 2025). Each dimension captures a critical aspect of AI-driven KM. Organizations assess each area using mixed methods (surveys, usage logs, expert review) and score against targets or external benchmarks. Table 1 summarizes the dimensions with their evaluation focus and example metrics. After the table, we briefly explain each dimension and its indicators.

Table 1: Dimensions of the AI-KM benchmarking framework, with their focus areas and illustrative metrics

Dimension	Evaluation Focus	Example Indicators / Metrics
Governance & Strategy	Leadership support, ethical policies, strategic alignment	Presence of an AI-KM steering committee or oversight board Existence of formal AI ethics/compliance policies Alignment of AI-KM projects with business goals (e.g. % of projects with documented objectives)
Data & Content Quality	Accuracy and completeness of knowledge assets	% of content reviewed/updated annually Taxonomy consistency score (e.g. % of documents correctly tagged Data error rate or duplicate content percentage (avoiding "data chaos")
Technology & Integration	AI model performance and system reliability	AI answer accuracy (%) on test queries Average response latency (seconds) System uptime (%) and error rates Ability to handle multilingual or evolving data inputs
People & Culture	User adoption, collaboration norms	User satisfaction rating (e.g. via survey) % of employees actively using the AI-KM tool KM training completion rate Participation in knowledge-sharing activities (e.g. contributions of new documents)
Performance Outcomes	Business impact and value creation	Average time-to-answer (e.g. helpdesk query resolution) Helpdesk ticket deflection rate (self-service usage) % Cost savings or ROI attributable to KM Number of new innovations or solutions generated using KM insights

In this framework, Governance & Strategy ensures AI-KM has executive backing and clear policies. For example, Nguyen et al. (2025) emphasize ethical oversight and clear guidelines. Data & Content Quality tackles the "Data Chaos" loop by enforcing currency and accuracy of content (e.g. 90% of documents updated yearly). Technology & Integration covers technical KPIs (e.g. ≥90% accuracy, <2s latency) and resilience. People & Culture measures adoption and trust – Paik & Cjnoi (2005) show that culture and training (e.g. ≥80% staff training completion) are critical for KM success. Performance Outcomes tracks concrete benefits: for instance, goals might include cutting average resolution time to under 2 minutes or deflecting ≥30% of support tickets. Accenture's experience – achieving approximately 60–65 percent time savings on a 100,000-hour authoring task – provides a target benchmark for ROI.

By collecting data through system logs, user surveys, and expert audits (following the triangulation approach described in Section 3), organisations can score each dimension against these targets. This balanced scorecard highlights weak areas (e.g. low user adoption or data quality issues) and guides continuous improvement. In sum, this multidimensional benchmarking embeds continuous evaluation into the KM lifecycle, ensuring that AI augments knowledge work sustainably.

6. Future Research

Based on our findings, we outline several directions for further study:

Privacy-Preserving AI Methods: Both Gelashvili-Luik et al. (2025) and Nguyen et al. (2025) identify data privacy as critical. Future work should develop and validate techniques (e.g. advanced anonymization, federated learning) that allow AI models to train on knowledge data without exposing sensitive information. Empirical studies could test how such techniques affect AI-KM utility and compliance with regulations.

Participatory and Trustworthy AI: Gelashvili-Luik et al. (2025) emphasize stakeholder participation in AI development. Research could explore co-design processes for KM bots, ensuring end-users contribute to model

training. For example, involving KM experts in labelling or validation could improve answer credibility. Studies should also investigate how transparency tools (e.g. explanations of AI answers) affect user trust and system adoption.

Ethical Frameworks for AI-KM: There is a need for comprehensive guidelines addressing bias, fairness, and accountability in AI-KM. Nguyen et al. (2025) suggest embedding digital ethics into all KM processes. Future work should define ethical KPIs for AI-KM and develop auditing methods (e.g. bias detection in knowledge bases) to ensure ethical alignment. Case studies in regulated sectors (healthcare, finance) would be valuable to understand domain-specific issues.

Tacit Knowledge Integration: Given the gap in capturing tacit expertise, research should investigate methods to surface such knowledge. This could involve AI-driven social network analysis to identify expert communities, or tools that prompt employees to share contextual insights. Experimental evaluations could measure how much additional value is gained by incorporating tacit knowledge into AI-KM.

Empirical Validation of Benchmarks: Our proposed framework needs testing in real organizations. Future research should apply it in case studies across industries, collecting data and comparing framework scores with organizational outcomes. This will refine which metrics are most predictive of success. In particular, longitudinal studies could link improvement in benchmark dimensions to actual business value from AI-KM.

These avenues will deepen understanding of AI in KM and help practitioners address the complex, socio-technical challenges involved.

7. Conclusion

AI-driven systems have the potential to greatly enhance enterprise knowledge management, as shown by Accenture's Knowledge Assist. However, realizing this potential requires attention to more than just technology. Our research shows that data quality, user trust, organizational culture, and ethics are equally important for successful AI-KM (Gelashvili-Luik et al. 2025; Rezaei 2025). The case of Accenture highlights both the promise (efficiency gains) and current gaps (governance, tacit knowledge). The benchmarking framework proposed here offers a holistic way to assess AI-KM maturity across multiple dimensions, guiding firms to strengthen their weak areas. By using this framework, organizations can iteratively improve their KM processes – ensuring that AI contributes to knowledge sharing and innovation in a measured, responsible manner.

Ethics Declaration: This research did not involve human subjects, personal data, or experimental interventions; it is based solely on analysis of publicly available documents and published literature. Therefore, ethical clearance was not required.

AI Declaration: We confirm that generative AI tools (OpenAI's ChatGPT) were used to assist in literature review and drafting of sections. All AI-generated content was critically reviewed, edited, and approved by the authors to ensure accuracy and adherence to scholarly standards.

References

Alavi, M. and Leidner, D.E. (2001) "Knowledge management and knowledge management systems: conceptual foundations and research issues", *MIS Quarterly*, 25(1), pp.107–136. doi:10.2307/3250961.

Braun, V. and Clarke, V. (2006) "Using thematic analysis in psychology", Qualitative Research in Psychology, 3(2), pp.77–101. doi:10.1191/1478088706qp063oa.

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12. https://doi.org/10.1177/2053951715622512

Cai, C. J., Winter, S., Steiner, D., Wilcox, L., & Terry, M. (2019). "Hello AI": Unpacking the black box of AI in healthcare. *ACM Transactions on Computer-Human Interaction, 26*(2), Article 12, 1–31. https://doi.org/10.1145/3313831

Coveo (2023) Knowledge Management Maturity: Benchmarks and Best Practices. Available at: https://www.coveo.com/en/resources/ebooks/knowledge-management-maturity (Accessed: April 2026).

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review, 96*(1), 108–116.

Gelashvili-Luik, T., Vihma, P. and Pappel, I. (2025) "Navigating the AI revolution: challenges and opportunities for integrating emerging technologies into knowledge management systems – a systematic literature review", *Frontiers in Artificial Intelligence*, 8:1595930. doi:10.3389/frai.2025.1595930.

Geller, I., Yang, S., Pillai, J., Kwatra, S. and Sonti, K. (2023) 'Accenture creates a Knowledge Assist solution using generative AI services on AWS', *AWS Machine Learning Blog*, 28 September. Available at: https://aws.amazon.com/blogs/machine-learning/accenture-creates-a-knowledge-assist-solution-using-generative-ai-services-on-aws/ (Accessed: April 2026).

Minonne, C. and Turner, G. (2018) "Measuring knowledge management performance: a framework and case study", *Knowledge Management Research & Practice*, 16(1), pp.5–19. doi:10.1080/14778238.2017.1367958.

Nguyen, L.T.K., Connolly, J. and Nguyen, H.N. (2025) "A systematic review of improving knowledge management with generative AI and large language models", *Journal of Advances in Information Technology*, 16(4), pp.594–612. doi:10.12720/jait.16.4.594-612.

Paik, Y. and Cjnoi, D.Y. (2005) "The shortcomings of a standardized global knowledge management system: The case study of Accenture", *Academy of Management Executive*, 19(2), pp.98–112. doi:10.5465/AME.2005.16963096.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Rezaei, M. (2025) "Artificial intelligence in knowledge management: identifying and addressing the key implementation challenges", *Technological Forecasting and Social Change*, 217, p.124183. doi:10.1016/j.techfore.2025.124183.

Szydełko, M. and Szydełko, Ł. (2025) "Informational support for benchmarking in the context of knowledge management", *European Research Studies Journal*, 28(3), pp.32–48. doi:10.35808/ersj/4031.

TSIA (2023) Technology and Service 3.0: The TSIA Knowledge Management Framework. Technology & Services Industry Association. Available at: https://www.tsia.com (Accessed: April 2026).