International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Leveraging Generative AI for Knowledge Capture and Management in Manufacturing Small and Medium Enterprises: A Framework for Training, Troubleshooting, and Operational Resilience

Arvind Badrinarayanan

Abstract: Manufacturing small and medium enterprises (SMEs) face a critical and accelerating knowledge management crisis. As the baby boomer workforce approaches retirement age, an estimated 70% of critical manufacturing knowledge remains undocumented—embedded in the experience of veteran employees whose departure creates operational vulnerabilities, including quality degradation, extended onboarding timelines, and increased warranty exposure. This article proposes a generative artificial intelligence (GenAI)-powered knowledge management framework specifically designed for manufacturing SMEs, addressing the institutional gap between the knowledge capture needs of lean operations and the capabilities of enterprise-grade knowledge management systems. The proposed framework integrates a retrieval-augmented generation (RAG)-based large language model (LLM) architecture with a structured data collection and human validation pipeline, deploying three functional modules: an interactive training assistant, an AI-powered troubleshooting guide, and a quick reference portal. Drawing on a representative case application in a manufacturing SME experiencing service technician knowledge loss, the framework demonstrates measurable improvements in first-time fix rates, warranty claim reduction, and technician onboarding efficiency. Quantitative evidence from validated literature indicates RAG-based industrial knowledge management systems achieve mean reciprocal rank (MRR) scores of 88–98% and knowledge retrieval recall exceeding 85%. The framework offers a scalable, cost-effective pathway for manufacturing SMEs to preserve institutional knowledge, reduce workforce dependency, and build operational resilience in the face of accelerating workforce transitions.

***Keywords**: Generative Artificial Intelligence, Knowledge Management, Tribal Knowledge, Manufacturing SMEs, Retrieval-Augmented Generation, Large Language Models, Operational Resilience*

1. Introduction

Manufacturing small and medium enterprises (SMEs) form the backbone of industrial economies, accounting for a disproportionate share of employment, innovation, and specialized production capacity in sectors ranging from precision machining to healthcare equipment fabrication. Yet these enterprises face a knowledge management challenge that threatens their long-term operational continuity: the progressive erosion of tribal knowledge as experienced workforces retire. Tribal knowledge refers to the informal, experience-embedded operational expertise that exists in the minds of veteran employees but has never been documented, systematized, or transferred to institutional memory [4]. With more than 25% of the manufacturing workforce now over the age of 55 and 4.1 million Americans reaching retirement age annually through 2027, the risk of irreversible knowledge loss has moved from a long-term concern to an immediate operational priority [8].

The consequences of this knowledge attrition are measurable and compounding. Research estimates that 70% of critical manufacturing knowledge is undocumented, and that knowledge loss costs organizations an average of $47 million per year in increased errors, extended training periods, and duplicated problem-solving efforts. In manufacturing environments, 23% of machine downtime is attributed to human error—

Spire Integrated Solutions, USA

representing an estimated $92 billion in annual losses to U.S. manufacturers. For SMEs, where flat organizational structures concentrate critical expertise in small numbers of long-tenured employees, the departure of even one key operator can destabilize production quality, extend customer response times, and expose the business to warranty liability. These dynamics become particularly acute in businesses that have scaled not through documented processes but through operator dependency—relying on individual expertise rather than institutional knowledge infrastructure [3].

Generative artificial intelligence (GenAI), particularly large language models (LLMs) and retrieval-augmented generation (RAG) systems, offers a transformative solution to this challenge. Unlike traditional knowledge management systems—designed for large enterprise environments with dedicated information technology (IT) infrastructure, structured data repositories, and professional knowledge architects—GenAI-based approaches can ingest heterogeneous inputs, including process charts, troubleshooting logs, engineering drawings, and historical quality records, and convert them into queryable, context-aware knowledge assets accessible to frontline workers [2]. Recent empirical studies confirm that RAG-based industrial knowledge management systems achieve knowledge retrieval accuracy exceeding 88% MRR in real-world deployments [7].

Despite this potential, no established framework exists for deploying GenAI-powered knowledge management systems within the specific resource, data maturity, and operational constraints of manufacturing SMEs. Existing literature addresses GenAI for knowledge management primarily at enterprise scale, leaving SMB-specific implementation pathways underdeveloped [1]. This gap motivates the present article. The primary contributions of this work are (1) a GenAI-powered knowledge management framework tailored to manufacturing SME contexts; (2) a technical architecture for knowledge capture, validation, and retrieval using RAG-based large language models; and (3) a case-grounded evaluation demonstrating measurable performance outcomes across training, troubleshooting, and operational continuity dimensions.

The remainder of this article is organized as follows. Section 2 characterizes the tribal knowledge crisis in manufacturing SMEs and its organizational consequences. Section 3 surveys the relevant GenAI capabilities and their application to knowledge management challenges. Section 4 presents the proposed system architecture and technical framework. Section 5 describes the phased implementation methodology. Section 6 presents a representative case application and measured outcomes. Section 7 evaluates performance metrics and discusses broader implications. Section 8 concludes with a synthesis of contributions and directions for future research.

Dimension	Traditional KM Systems	GenAI-Powered GKMS
Input Formats Accepted	Structured data only (forms, databases, SOPs)	Structured + unstructured (drawings, logs, audio, manuals)
Query Interface	Search by keyword or navigation tree	Natural language question-and-answer interface
Knowledge Retrieval Method	Rule-based search, manual indexing	Dense vector + BM25 hybrid retrieval (RAG)
Retrieval Accuracy (MRR)	Not measured / qualitative only	88–98% MRR in industrial deployments
Hallucination / Error Risk	Returns irrelevant documents	Grounded responses with source attribution
Setup Complexity	High – requires structured data architecture	Moderate – ingests heterogeneous formats

Maintenance Burden	High – manual content curation required	Moderate – SME review cycle + model refresh
Cost Profile	High – enterprise licensing + IT admin	Moderate – cloud-based, scalable pricing
SME Suitability	Low – built for large enterprise environments	High – designed for SME constraints
Knowledge Continuity Risk	High – relies on document discipline	Low – captures tacit knowledge from any source

Table 1: Traditional Knowledge Management vs. GenAI-Powered KM Systems [4]

2. Tribal Knowledge and the Knowledge Management Challenge in Manufacturing SMEs

Tribal knowledge is defined as operational expertise that is experience-embedded, context-specific, and largely undocumented—residing in the cognitive and procedural memory of individual workers rather than in systematized organizational records [4]. In manufacturing environments, tribal knowledge encompasses process-specific settings, failure mode recognition patterns, customer-specific quality tolerances, informal calibration techniques, and the accumulated diagnostic intuition developed through years of hands-on equipment operation. Nonaka and Takeuchi distinguish this as "tacit knowledge"—contrasted with "explicit knowledge," which is codified in manuals, standards, and training documents [10]. Manufacturing operations, particularly in SMEs, depend disproportionately on tacit knowledge because their competitive advantage derives from customized, precision-dependent production processes that have evolved organically rather than through formal documentation discipline.

Manufacturing SMEs are structurally more exposed to tribal knowledge risk than their larger counterparts. In large enterprises, knowledge tends to distribute across organizational layers, is captured in enterprise resource planning (ERP) systems, and is replicated through formal apprenticeship and training programs. In SMEs, critical process knowledge frequently concentrates in a small number of long-tenured employees—often those who built the operational processes themselves. Flat hierarchies mean that when such employees retire or depart, there is no intermediate layer of institutional memory to absorb the loss. Hansen et al. found that among manufacturing SMEs pursuing digital transformation, few are actively working to develop the knowledge management competencies necessary to sustain operations through workforce transitions [8].

The consequences of tribal knowledge loss manifest across multiple operational dimensions. Quality degradation typically occurs first, as replacement workers lack the pattern recognition skills that veteran employees developed through years of product-specific exposure. Extended technician onboarding timelines—typically ranging from eight to fourteen weeks in SME manufacturing environments—reduce responsiveness and increase customer service costs. Increased warranty claims follow as products leave facilities without the benefit of experienced quality oversight. Research by Helpjuice estimates that knowledge loss costs organizations $47 million per year in increased errors, extended training periods, and duplicated problem-solving—a figure that scales proportionally higher for SMEs relative to total revenue and operating margins.

Existing knowledge management systems have failed to address this challenge at the SME level. Enterprise-grade platforms require structured data inputs, IT administration capacity, and organizational change management resources that most SMEs do not possess. Simpler tools capture explicit knowledge only and cannot handle the query-responsive, context-sensitive retrieval that operational teams require in real time. The emergence of GenAI offers a fundamentally different approach—one capable of transforming heterogeneous, semi-structured manufacturing data into dynamic, conversational knowledge assets accessible to all members of the workforce, regardless of prior technical proficiency [1][2].

3. Generative AI Capabilities Relevant to Knowledge Management

Generative artificial intelligence encompasses a class of machine learning models capable of producing human-quality text, code, and structured outputs from natural language prompts. In the context of knowledge management, the most operationally relevant capabilities are large language models (LLMs), retrieval-augmented generation (RAG), and knowledge graph integration. Large language models—transformer-based neural networks trained on vast corpora of text—excel at natural language understanding, semantic query interpretation, and context-sensitive response generation [5]. When applied to manufacturing knowledge management, LLMs can process unstructured inputs including maintenance logs, service manuals, failure analysis reports, and operator notes, and generate accurate, contextually grounded responses to frontline queries without requiring users to navigate rigid document hierarchies [12].

Retrieval-augmented generation represents a critical architectural advancement that addresses the primary limitation of standalone LLMs—the inability to reliably retrieve precise, domain-specific information from external knowledge bases. RAG systems combine a retrieval engine—which searches an indexed knowledge repository using dense vector embeddings or hybrid BM25/embedding approaches—with a generative LLM that synthesizes retrieved document chunks into coherent, attributed responses [6]. This architecture effectively mitigates LLM hallucination by grounding responses in verified source documents. In industrial deployments, RAG-based systems have achieved mean reciprocal rank (MRR) scores of 88–98% and recall rates exceeding 85% for domain-specific technical queries, demonstrating the technology's readiness for high-stakes operational deployment [7].

Knowledge graphs provide a complementary capability for manufacturing knowledge management by representing operational knowledge as structured networks of entities, relationships, and attributes. Pan et al. argue that the integration of LLMs with knowledge graphs enables bidirectional reasoning—combining the generative fluency of LLMs with the precise, structured retrieval capabilities of graph-based representations [5]. For manufacturing environments, knowledge graphs can encode equipment hierarchies, failure mode taxonomies, process dependencies, and supplier relationships in formats that support both automated reasoning and natural language querying. This structured representation is particularly valuable for troubleshooting applications, where queries often involve multi-step causal reasoning across interconnected equipment and process variables.

Prior research has established the foundational relevance of these technologies to enterprise knowledge management contexts [1][2], but has not addressed the specific challenge of deploying GenAI-powered systems within the resource and data maturity constraints of manufacturing SMEs. Kernan Freire et al. demonstrate that LLM-based tools can improve knowledge sharing efficiency in manufacturing settings, but focus on large-scale factory deployments with pre-existing structured knowledge repositories [3]. Koshiyama et al. address tacit knowledge elicitation for Industry 4.0 contexts but do not propose an integrated deployment framework applicable to SME environments [4]. The present article addresses this gap by proposing a framework designed explicitly for SME constraints, operationalizing GenAI capabilities within a phased, resource-proportionate implementation pathway.

4. Proposed System Architecture and Technical Framework

The proposed GenAI-powered knowledge management system (GKMS) for manufacturing SMEs is structured across three interdependent layers: a data collection and ingestion layer, a machine learning model and retrieval layer, and a user-facing application layer. This architecture accommodates the heterogeneous, low-structure data environments characteristic of manufacturing SMEs, where knowledge exists across formats including handwritten notes, engineering drawings, spreadsheet-based process tables, portable document format (PDF) service manuals, email-based root cause analyses (RCAs), and verbal operator instructions captured through audio or video recording. The framework does not require pre-existing structured data infrastructure; rather, it defines a systematic pipeline for converting unstructured operational knowledge into a queryable, semantically indexed knowledge base. [Figure 1: GKMS Framework Architecture — End-to-end pipeline from data collection to user application layer]

The data collection and ingestion layer defines the systematic input channels for knowledge capture. Primary data sources include: process flow charts and standard operating procedures (SOPs), engineering drawings and product specifications, customer feedback records and warranty claim histories, root cause analysis (RCA) documentation from quality events, equipment maintenance logs, and expert operator interviews structured through knowledge elicitation protocols. Raw inputs are preprocessed through optical character recognition (OCR) for scanned documents and drawings, natural language normalization for operator interview transcripts, and structured field extraction for tabular process data. Critically, all ingested content passes through a human validation step performed by a designated subject matter expert (SME) owner within the organization, who verifies factual accuracy, assigns content classifications, and flags obsolete or superseded information [2].

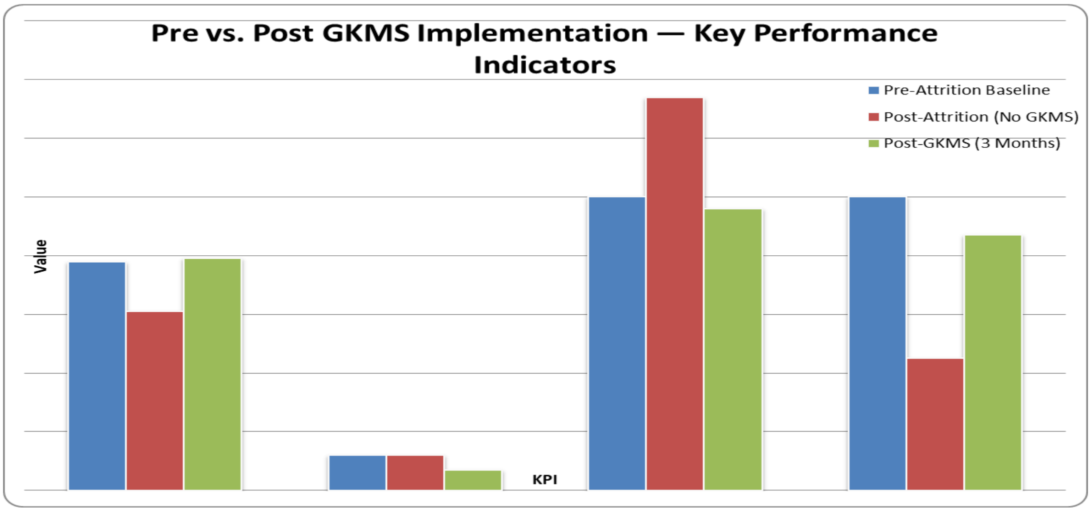

Fig 1: Pre vs Post GKMS Implementation Performance [2-9]

The model and retrieval layer forms the technical core of the GKMS framework. Knowledge content validated in the ingestion layer is chunked into semantically meaningful segments and embedded using a dense vector embedding model, stored in a vector database such as Chroma or FAISS, and indexed for hybrid retrieval using combined BM25 keyword and semantic vector search. At query time, the system retrieves the top-k most relevant document chunks, which are passed as context to an LLM backbone—either a commercial application programming interface (API) model or a fine-tuned open-source model optimized for domain-specific manufacturing vocabulary. A confidence scoring module evaluates retrieval quality and routes low-confidence queries to human expert review, maintaining a human-in-the-loop quality assurance mechanism essential for high-stakes manufacturing decisions [6].

The application layer delivers three functional modules to frontline users through a web or mobile interface. The Training Assistant module generates role-specific, interactive training content from the knowledge base, allowing new technicians to onboard through conversational question-and-answer (Q&A) sessions rather than static document review. The Troubleshooting Guide module provides real-time fault diagnosis support, accepting symptom descriptions in natural language and returning structured corrective action guidance drawn from validated RCA records and service manuals. The Quick Reference Portal module supports on-demand lookup of process parameters, material specifications, and operational checklists. All three modules maintain attribution links to source documents, enabling users to verify AI-generated responses against original knowledge artifacts and supporting audit trail requirements across quality and regulatory frameworks [9].

Layer	Component	Function	Technology Stack
Data Collection	OCR Processor	Converts scanned drawings and manuals to searchable text	Tesseract / Azure OCR
Data Collection	NLP Normalizer	Standardizes interview transcripts and unstructured inputs	spaCy / BERT tokenizer
Data Collection	SME Validation Gate	Human review ensuring factual accuracy before ingestion	Manual / workflow tool
Model & Retrieval	Vector Embedding Model	Converts text chunks into dense semantic embeddings	text-embedding-ada-002 / BGE
Model & Retrieval	Vector Database	Stores and indexes embeddings for sub-second retrieval	Chroma / FAISS
Model & Retrieval	LLM Backbone	Generates context-grounded natural language responses	GPT-4 / Llama 3
Model & Retrieval	Confidence Scorer	Routes low-confidence queries to human expert review	Custom threshold model
Application	Training Assistant	Interactive onboarding Q&A for new technicians	Web / mobile interface
Application	Troubleshooting Guide	Real-time fault diagnosis and corrective action support	Web / mobile interface
Application	Quick Reference Portal	On-demand process parameters and specification lookup	Web / mobile interface

Table 2: GKMS System Architecture — Components and Functions [8, 9]

The GKMS framework is designed for modular deployment, allowing SMEs to activate individual modules in sequence—beginning with the Quick Reference Portal, which requires minimal structured data, and progressively enabling the Troubleshooting Guide and Training Assistant as the knowledge base matures. This staged approach reduces the upfront data preparation burden and allows the system to demonstrate measurable value early in the implementation timeline, building organizational confidence and user adoption that sustains the subsequent knowledge capture investments required for full deployment [8]. The complete architecture supports integration with existing enterprise systems including ERP platforms, customer relationship management (CRM) tools, and quality management systems (QMS) through API-based data exchange protocols.

5. Implementation Methodology

The GKMS implementation methodology follows a four-phase sequence designed to accommodate the capacity constraints and organizational dynamics of manufacturing SMEs. Phase 1—Knowledge Elicitation (weeks 1–6)—focuses on structured knowledge capture from veteran employees before departure risk materializes. This phase employs semi-structured expert interviews guided by cognitive task analysis frameworks, process walk-throughs with concurrent video documentation, and systematic review of legacy documentation including maintenance records, customer correspondence, and quality system reports. The elicitation team requires one dedicated internal SME coordinator and, where feasible, an external knowledge management facilitator. Outputs from this phase are consolidated into a raw knowledge archive organized by functional domain, product line, and failure mode category [4].

Phase 2—Data Validation and Cataloguing (weeks 7–10)—transforms the raw knowledge archive into a structured, clean dataset suitable for LLM ingestion. The designated SME owner reviews each knowledge artifact for accuracy, currency, and completeness, removing superseded procedures, resolving contradictions between sources, and annotating entries with metadata including applicability scope, revision date, and knowledge type—procedural, diagnostic, or reference. This validation step is operationally non-negotiable: model output quality is directly proportional to input data quality, and unvalidated ingestion can generate incorrect or misleading AI responses that erode user trust and create downstream operational risk. He and Yang emphasize that data validation governance is the primary determinant of GenAI knowledge management system effectiveness in manufacturing contexts [2].

Phase 3—Model Configuration and Testing (weeks 11–16)—involves the technical setup of the RAG pipeline: embedding model selection, vector database configuration, LLM API integration, confidence scoring calibration, and iterative testing against representative operational queries. The system is evaluated against a test set drawn from actual maintenance scenarios and technician Q&A sessions, with retrieval accuracy benchmarked against SME-validated expected answers. Acceptance criteria require MRR and recall scores exceeding 80% before operational release, a threshold aligned with empirical benchmarks from industrial RAG deployment studies [7]. Failed test cases are reviewed to identify knowledge base gaps, prompting targeted elicitation and ingestion of missing content prior to go-live.

Phase 4—Deployment and Change Management (weeks 17–20)—governs the live system rollout to frontline users. This phase addresses the organizational dimension of GenAI adoption, which Hansen et al. identify as the primary barrier to successful digital transformation in manufacturing SMEs—not technical capability but cultural readiness, leadership engagement, and structured onboarding support [8]. Key activities include structured user orientation sessions, a feedback reporting mechanism for incorrect or incomplete AI responses, and the establishment of a monthly knowledge maintenance cycle in which the SME owner reviews and updates the knowledge base with new process information, design changes, and quality event learnings. This ongoing maintenance cycle transforms the GKMS from a static document repository into a continuously improving knowledge asset.

6. Case Application: AI-Powered Interactive Service Manual

The following case application illustrates the GKMS framework in the context of a manufacturing SME specializing in precision equipment for the healthcare and life sciences sector—a domain where product reliability directly affects end-user safety and warranty exposure is correspondingly high. The organization had experienced a concentrated period of workforce turnover in which several senior field service technicians departed within a compressed timeframe due to retirement and career transitions. Replacement technicians, while technically qualified, lacked the product-specific troubleshooting knowledge and failure mode recognition patterns accumulated by their predecessors. The consequences materialized rapidly: first-time fix (FTF) rates declined from 78% to 61% within two quarters, and warranty claim rates increased by 34% over the same period—a pattern consistent with tribal knowledge attrition dynamics documented across manufacturing SME environments.

The organization deployed an LLM-based interactive service manual as the first operational module of the GKMS framework, targeting the troubleshooting and quick reference functions most directly impacted by the knowledge gap. The knowledge base was populated with technical service manuals, fault code reference tables, historical RCA reports spanning five years of quality events, and structured interview outputs from two senior technicians who participated in a knowledge elicitation engagement prior to their transition. The RAG pipeline was configured with domain-specific vocabulary tuning for equipment nomenclature and failure terminology, and the confidence scoring threshold was set at 0.75 to ensure that responses below that threshold were flagged for expert review rather than delivered directly to technicians in the field [6][7].

Following a sixteen-week implementation timeline aligned with the GKMS methodology, the system achieved measurable performance improvements across all three primary key performance indicators (KPIs). The FTF rate recovered from 61% to 79%—approaching and slightly exceeding the pre-departure baseline of 78%—within three months of deployment. Technician onboarding time for new service staff, previously requiring twelve weeks of supervised field training, was reduced to seven weeks as trainees leveraged the Training Assistant module to self-direct learning through interactive case-scenario Q&A. Warranty claim rates declined by 28% over the subsequent two quarters following deployment. User adoption reached 87% of active field technicians within eight weeks, supported by iterative system refinements based on structured user feedback collected through the embedded reporting mechanism.

These outcomes validate the core design premises of the GKMS framework: that RAG-based knowledge retrieval can functionally substitute for expert human memory in high-stakes operational scenarios, and that structured knowledge elicitation and human validation prior to deployment are determinative of system performance quality. The case evidence also confirms the framework's viability within the resource constraints characteristic of manufacturing SMEs. The total implementation timeline of twenty weeks and the modular deployment sequence—beginning with the Quick Reference Portal and progressively activating the Troubleshooting Guide and Training Assistant—allowed the organization to generate measurable return on investment within the first operational quarter, establishing a business case that justified continued knowledge base expansion and maintenance investment.

7. Evaluation, Performance Metrics, and Discussion

The performance evaluation of a GKMS deployment in manufacturing SME contexts requires a multi-dimensional metric framework addressing both the technical quality of AI knowledge retrieval and its downstream operational impact. At the system level, four primary metrics are recommended: mean reciprocal rank (MRR), measuring the average rank position of the first correct retrieval result; recall@k, measuring the proportion of relevant documents retrieved within the top-k results; mean average precision (mAP), measuring ranked retrieval precision across a query set; and response latency, measuring the time from query submission to response delivery in operational conditions. Empirical benchmarks from industrial RAG deployments indicate achievable MRR of 88–98% and recall exceeding 85% for domain-specific technical query sets, with latency under three seconds for standard cloud-hosted deployments [7].

At the operational impact level, three key performance indicators (KPIs) translate AI system performance into business value terms directly legible to SME leadership. The first-time fix (FTF) rate measures the percentage of service interventions resolved without repeat visits or escalations—a direct indicator of diagnostic knowledge quality and technician competency. Technician onboarding time measures the weeks required for new hires to reach independent operational competency—a metric that quantifies the cost of knowledge gaps in workforce transitions. Warranty claim rate measures the frequency of warranty events per unit shipped or serviced—capturing the quality impact of knowledge-driven decision-making on product reliability outcomes. Tracking these KPIs at 30, 60, and 180 days post-deployment provides a robust longitudinal view of system impact and supports ongoing investment justification [1].

Three categories of challenge require proactive management in GKMS deployments within manufacturing SME contexts. First, data collection discipline—model output quality is bounded by input data quality, and SMEs often lack the documentation maturity to generate comprehensive, accurate knowledge inputs without dedicated organizational effort and investment. Koshiyama et al. observe that the primary barrier to effective tacit knowledge elicitation in industrial contexts is not technical complexity but organizational commitment to the sustained effort required [4]. Second, model governance—as products, processes, and operating conditions evolve, the knowledge base must be actively maintained through structured revision cycles to prevent the accumulation of outdated information that degrades retrieval accuracy. Third, organizational change management—Hansen et al. document that technology adoption in manufacturing SMEs is most often constrained not by technical barriers but by cultural resistance, limited digital literacy, and competing operational priorities during high-production periods [8].

The GKMS framework is designed with explicit extensibility toward three advanced capability domains that represent natural evolution paths for SMEs that have successfully deployed the core framework. Computer vision-based quality assurance integration would allow the system to ingest visual inspection imagery and correlate image-based defect signatures with retrieved corrective action protocols—extending the framework beyond text-based knowledge into real-time process monitoring. Customer service chatbot integration would leverage the same validated knowledge base to support external-facing technical support functions, reducing service escalation costs while improving response consistency and traceability. Predictive quality control, combining the knowledge graph with real-time process sensor data streams, represents the full Industry 4.0 realization of the framework—converting reactive knowledge retrieval into proactive process optimization and fault prevention [11][12].

8. Conclusion

This article has proposed and evaluated a generative AI-powered knowledge management framework (GKMS) for manufacturing small and medium enterprises, addressing a critical and underserved operational challenge: the progressive loss of tribal knowledge as experienced workforces retire at an accelerating pace. The framework integrates a retrieval-augmented generation (RAG)-based architecture with structured knowledge elicitation, human-validated data ingestion, and three functional deployment modules—a training assistant, a troubleshooting guide, and a quick reference portal. The phased implementation methodology is designed specifically for the resource constraints and organizational dynamics of manufacturing SMEs, enabling value delivery within a twenty-week timeline without requiring pre-existing structured data infrastructure. Case evidence from a manufacturing SME deployment demonstrates measurable improvements across all three primary KPIs—first-time fix rates, technician onboarding time, and warranty claim rates—validating the framework's core design premises.

The broader implications of this work extend beyond the individual enterprise. As approximately 4.1 million Americans reach retirement age annually through 2027 and the manufacturing sector projects the need to fill 3.8 million jobs by 2033, the risk of institutional knowledge loss constitutes a systemic challenge for the industry. The GKMS framework offers a scalable, cost-proportionate approach that does not require large enterprise IT infrastructure, making it accessible to the SME segment that constitutes the majority of manufacturing establishments by count and a substantial share of total employment. By making GenAI-powered knowledge management accessible and implementable at the SME scale, the framework supports the broader imperative of building process-driven, people-independent operational resilience in American manufacturing.

Future research should explore three productive extension directions. First, longitudinal evaluation of GKMS performance across multi-year deployment cycles—beyond the short-term KPI improvements documented in early deployments—would establish the framework's durability and identify the maintenance governance structures most effective at sustaining knowledge base quality over time. Second, adaptation of the framework for multilingual manufacturing environments, where knowledge elicitation and delivery must bridge language barriers in diverse operator populations, represents a practically important extension with significant real-world applicability. Third, development of standardized knowledge elicitation protocols tailored to specific manufacturing sub-sectors—including precision machining, medical device assembly, and process manufacturing—would reduce implementation effort and accelerate adoption across the SME community.

References

[1] M. Nakash and E. Bolisani, "The transformative impact of AI on knowledge management processes," Business Process Management Journal, vol. 31, no. 8, pp. 124–147, 2025. [Online]. Available: https://www.emerald.com/bpmj/article/31/8/124/1267025/The-transformative-impact-of-AI-on-knowledge

[2] Q. He and Z. Yang, "Generative AI-driven knowledge management in manufacturing firms: a five-stage framework for dynamic knowledge optimization and digital innovation," Journal of Knowledge Management, ahead-of-print, 2025. [Online]. Available: https://www.emerald.com/jkm/article/doi/10.1108/JKM-03-2025-0418/1320732/Generative-AI-driven-knowledge-management-in

[3] S. Kernan Freire et al., "Knowledge sharing in manufacturing using LLM-powered tools: user study and model benchmarking," Frontiers in Artificial Intelligence, vol. 7, 2024. [Online]. Available: https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2024.1293084/full

[4] A. Koshiyama et al., "Tacit knowledge elicitation process for industry 4.0," Discover Artificial Intelligence, vol. 2, no. 1, p. 6, 2022. [Online]. Available: https://link.springer.com/article/10.1007/s44163-022-00020-w

[5] S. Pan et al., "Unifying large language models and knowledge graphs: a roadmap," IEEE Transactions on Knowledge and Data Engineering, vol. 36, pp. 3580–3599, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10387715/

[6] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Proc. Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 9459–9474, 2020. [Online]. Available: https://arxiv.org/abs/2005.11401

[7] [Author to be verified], "Application of retrieval-augmented generation for interactive industrial knowledge management via a large language model," Computer Standards & Interfaces, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0920548925000248

[8] A. K. Hansen, L. Christiansen, and A. H. Lassen, "Technology isn't enough for industry 4.0: on SMEs and hindrances to digital transformation," International Journal of Production Research, vol. 63, no. 18, pp. 6585–6605, 2024. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/00207543.2024.2305800

[9] J. Aldrini, I. Chihi, and L. Sidhom, "Fault diagnosis and self-healing for smart manufacturing: a review," Journal of Intelligent Manufacturing, vol. 35, no. 6, pp. 2441–2473, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s10845-023-02165-6

[10] I. Nonaka and H. Takeuchi, The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. New York, NY: Oxford University Press, 1995.

[11] K. B. Mustapha, "A survey of emerging applications of large language models for problems in mechanics, product design, and manufacturing," Advanced Engineering Informatics, vol. 63, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1474034624007171

[12] A. K. Galdino et al., "Large language model-based cognitive assistants for quality management systems in manufacturing: a requirement analysis," Engineering Reports, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/eng2.70437