International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group | Volume 4, Issue 4, 139-151, 2023
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V4I4P116

Original Article

AI-Driven Knowledge Management Systems for Enterprise IT Operations

Nareddy Abhireddy,
Independent Researcher, USA.

Abstract *- Artificial Intelligence (AI) is increasingly being integrated into Knowledge Management (KM) systems to provide enhanced decision-making support to Information Technology (IT) operations teams. AI-driven systems are empowering these teams to reduce service resolution times, streamline troubleshooting processes, and share knowledge more effectively by improving the speed, relevance, and contextual accuracy of knowledge retrieval and understanding. Alongside these advantages, such AI-powered capabilities are redesigning the nature of knowledge work, augmenting the collaborative aspect of KM efforts, and shifting the composition and skill requirements of IT operations teams. Fundamental considerations for developing AI-driven KM systems for IT operations are explored, including a discussion of deployed architectural patterns, data quality and governance concerns, and deployment and processing options. Further, the impact of these systems and capabilities on the KM aspect of IT operations teams is examined. Together, these insights provide a foundation for the adoption of AI capabilities within KM systems used by IT operations teams.*

Keywords *- Cross-Border E-Commerce Data Governance; Cloud Computing; Data Localization And Residency; Data Sovereignty; Service-Oriented Architecture; Scalable Architectures; Third-Party Risk Management; Trusted Identity Management; Unified-Consent Framework.*

1. Introduction

The increasing complexity of enterprise IT operational environments, the growing demand for high-quality service availability and performance, and the imperative for IT cost management have led to the outsourcing of many recurring operational processes. Yet, some IT services that are neither core business services nor capable of being fully automated remain in-house. Examples include monitoring, incident detection and response, event management, service request fulfilment, and problem resolution. Knowledge management is crucial to these processes because, in ideal conditions, operations teams would efficiently solve issues by exploiting existing procedures, diagnostics, workarounds, or root-cause analyses. The ability to capture, curate, and distribute this knowledge among teams in a timely manner can reduce resolution time, improve service quality, and minimise the risk of problems.

Regardless of whether outsourced operations supply their services with or without knowledge, integrating all knowledge into a KM system helps operations and core teams scale effectively without sacrificing quality. An AI-driven KM system aligns with the trends of federated service management and product-led IT adoption. The deployment of artificial intelligence capabilities enriches existing KMS, enabling IT teams to supply and consume knowledge in a more intuitive way. These systems then reduce knowledge contribution/curation effort by capturing it directly from the conversations of team members, helping these KM solutions finally fulfil Agvent's promise of being a non-intrusive experience.

Fig 1: Generative AI in Knowledge Management

1.1. Background and Significance

Business organizations continuously strive to provide their products and services in an intelligent manner that combines speed with ingenuity. They have naturally gravitated toward the 21st-century business technology platform comprised of cloud computing, big data and analytics, mobile computing, social media, and the Internet of Things, which enable and encourage timely access to, and exchange of, their data and information in all forms. Ideally, such organizations would like to apply the same level of intelligence to all of their repeatable future modes of behavior, whether responding to market opportunities, detecting and preventing operational problems that could disrupt service delivery, performing operational functions like marketing and cross-selling, or maintaining their business technology platform not just to their production-level job execution but also to the detection of production problems and their root causes. Emulating their real-time detected detections of production problems in a collaborative manner involves designating knowledge workers (KWs) to work together, as a team, on new, similar problems. Likewise, the information action business organizations would like to experience is the residual benefit that can be obtained from analyzing problems after they have been resolved.

Equation 1: Derived equations (step-by-step) you can use for this topic
Let:

- Query text $q \rightarrow$ embedding vector $\mathbf{e}_q \in \mathbb{R}^d$
- Document $d_i \rightarrow$ embedding vector $\mathbf{e}_i \in \mathbb{R}^d$

Step 1: Dot product

$$\mathbf{e}_q \cdot \mathbf{e}_i = \sum_{j=1}^{d} (\mathbf{e}_q)_j (\mathbf{e}_i)_j$$

Step 2: Vector magnitudes

$$\| \mathbf{e}_q \| = \sqrt{\sum_{j=1}^{d} (\mathbf{e}_q)_j^2}, \| \mathbf{e}_i \| = \sqrt{\sum_{j=1}^{d} (\mathbf{e}_i)_j^2}$$

Step 3: Cosine similarity score

$$s(q,i) = \cos(\theta) = \frac{\mathbf{e}_q \cdot \mathbf{e}_i}{\| \mathbf{e}_q \| \ \| \mathbf{e}_i \|}$$

1.2. Research design

Knowledge management in enterprise IT operations prevention and resolution of service outages and incidents, and fulfillment of change requests is increasingly driven by artificial intelligence (AI). AI optimizes the retrieval of knowledge, improves understanding of context and intent, and adds value wherever knowledge is applied. AI performs exploratory analysis on bodies of knowledge and generates new knowledge elements. Knowledge management for enterprise IT operations requires specific considerations: patterns of system architecture, data governance and quality, operational deployment, and skill profiles.

Research on enterprise IT operational knowledge management KM in this work identifies patterns, noting that AI enables real-time operation in a cloud-native or cloud-like deployment model. IT operations personnel are relieved of some routine, knowledge-related tasks, but redesign of knowledge work and processes, and support for collaboration and skill evolution, remain priorities.

2. Foundations of Knowledge Management in IT Operations

Knowledge management (KM) is a sought-after objective in enterprise IT operations. Despite the proximity of the concepts, major KM frameworks lack integration with the business service-oriented context of IT operations and the specific functions of the IT operations team. Enterprise IT operations provide the knowledge work foundation. The Service Operation lifecycle offers the KM framework. IT operations teams produce and use a knowledge product.

The KM lifecycle in IT operations takes place at a Service Operations level in the IT Service Management Framework, a service-oriented adaptation of IT Infrastructure Library best practices. Each of the KM lifecycle phases identifies processes, roles, deliverables, and the data that management and knowledge operations teams transform into knowledge, which Knowledge Management is the part of the Operations Lifecycle that aims to ensure that the knowledge and information that are necessary for the effective delivery of Service Operation are available to the right people at the right time. From the other side, the evolution of enterprise IT has been characterized by the development of IT management capabilities that seeks to improve an enterprise's ability to manage how IT Systems are delivered, supported and operated for business services toward integration between business and IT.

Fig 2: Artificial Intelligence and Knowledge Management

2.1. Definitions and scope

Knowledge is critical to the delivery and effective support of IT services, yet knowledge work in IT operations is often poorly supported and can easily degrade in quality. Knowledge Management seeks to close this gap by establishing processes and systems that enable the effective storage, retrieval, and creation of knowledge, with the goal of improving quality and efficiency. This section discusses the nature of IT Operations knowledge work and provides a definition of knowledge management in enterprise IT Operations that accommodates these unique requirements.

Knowledge is defined here as actionable information that is important for executing the organization's mission. In the context of IT Operations, the mission is to deliver and support IT services by performing technical tasks – often called "IT Operations work" or simply "work" as efficiently as possible and with minimal adverse effect on the business operations supported. The successful execution of this work is typically supported by currently accepted procedures, backed by knowledge repositories or sources, together with practical experience and a thorough understanding of the IT environment. If the knowledge contained in these sources is insufficient to perform the work safely and efficiently, it is often enhanced through dialog with other personnel. Knowledge Management for IT Operations refers to "the processes and systems that support the creation, transfer, retention, and retrieval of knowledge and other relevant assets within and between IT Operations teams, to improve overall efficiency and quality of service delivery."

Equation 2: Precision@k, Recall@k (to measure "retrieval improvement")

Let:

- R_k= set of top-k retrieved docs
- G= set of truly relevant docs ("gold" relevance)

Precision@k

$$P@k = \frac{| R_k \cap G |}{k}$$

Recall@k

$$R@k = \frac{| R_k \cap G |}{| G |}$$

F1@k (optional combined score)

$$F1@k = \frac{2(P@k)(R@k)}{P@k + R@k}$$

2.2. Knowledge lifecycle in enterprise IT

Enterprise IT operations are knowledge-intensive. The complexity of enterprise architecture and the scale of operations create inherent monolithic challenges. Repeated occurrences of IT outages, with visible impact on business services, demonstrate the weakness of enterprise IT operations teams in extracting the knowledge of past incidents to prevent, or at least predict, such outages. This is not because knowledge is not captured or stored; it is because the evolutionary processes for knowledge management are not enabled with sufficient fluidity. KM is not just about zestfully managing knowledge. KM in enterprise IT operations is about capturing, curating, and publishing hot–cold knowledge consumable by people. It is about language–model-based generative AI systems composing the knowledge products to target the specific consumer in the specific context. Knowledge product developers are knowledge workers, much like authors or publishers, but operate only within the boundaries of an enterprise product catalog.

In a Knowledge Cycle, recipients of hot knowledge become its authors. When a hot knowledge product is consumed by a person who is planning a similar or different action, that action represents a query for knowledge retrieval. Based on feedback, the Hot Knowledge product can be bifurcated into two paths: a smaller set of actions deployed multiple times are candidates for Hot-Product-Cool-Knowledge; and newly constructed documentation groups and knowledge templates—like training and certification guides—when appropriately tagged become candidates for Hot-Product-Bat-Knowledge. On the longer time-scale, non-repeated actions leading to repeatable outcomes contribute to Hot-Product-Cool-Products, and Hot-Catalog-Cool-Catalog. Processes such as training and certification of IPs creating Cool-Knowledge-Product-Clusters and Knowledge-Product-Facilitators enable business user groups to efficiently manage Hot-Knowledge-Product-Cool-Knowledge and Hot-Knowledge-Consumer-Cool-Knowledge-Creation.

3. Role of Artificial Intelligence in Knowledge Management

Research on AI technologies, systems, and tools capable of augmenting the KML process re-examines the knowledge lifecycle and the KML systems architecture using news articles, press releases, speeches, and vendor materials from AI providers and organizations developing AI-driven KMS for enterprise IT operations. To understand how such systems enrich KMS for enterprise IT operations, two functions of AI are particularly relevant: retrieval-augmented generation capabilities that increase the KMS's retrieval capabilities and AI-powered NLP that enables automated understanding and contextualization of intents. Research findings point to a novel architectural pattern for KMS systems for IT operations, facilitating the offloading of both knowledge generation and KML curation processes. A knowledge worker-centric view points out the implications of an AI-driven KMS for IT operations teams and the IT operations function overall.

Several factors limit the ability of existing KMS to truly support a knowledge-centric approach for enterprise IT operations. First, despite a wealth of data contained in several IS components (e.g., application, service, incident, change, asset databases), the connections across these and other sources remain buried. Second, while KMS systems attempt to create connections among these data sources, efforts are often concentrated on a subset of IS components, reducing their context richness. Third, KMS are embedded in complex, integrated systems that cannot be easily queried by operations teams members. Consequently, KMS systems are treated as static repositories updated and enriched at irregular intervals that do not match enterprise KML needs.

Fig 3: Knowledge Management with AI Technology

3.1. Retrieval and search optimization

Rich knowledge repositories are essential to aid in the optimal distribution and utilization of digital knowledge across large, geographically dispersed organizations. Functionally similar groups often work in close proximity, dealing with similar issues and creating similar solutions. These groups generate substantial numbers of tickets and knowledge articles. Unequal knowledge quality across the lifecycle of the knowledge article results in varying resolution times, with the knowledge base serving more as a reference for culprits to locate a point in time when systems were in a healthier state than for storm management. Self-service portals reduce workload but can also mask Service Desk function deficits for business users. AI technologies can improve solutions for such organizations by reducing repetitiveness in work tasks—intelligent automation to sift through knowledge work—and by assisting in a better, more intuitive interface for engaging with knowledge bases. An AI engine operates in the cloud and learns from these knowledge-basis support tasks and articles, detects and fills in gaps, and has an improving effect for the remainder.

Intelligent retrieval systems save time when retrieving infrastructure knowledge, filtering past infrastructure states that had no exploratory interest to the requester. AI decisions can also optimize operational planning with intelligent function-discontinuation scheduling for testing and maintenance, creating a built-in knowledge creation plan. Machine learning (ML) algorithms for ticket closure prediction can assist system administrators in choosing the best runbook to minimize mean troubleshooting time and hence improve operational excellence. Advanced AI techniques can detect low-quality Web results for common queries at multiple levels of importance, optimizing user experience by interacting appropriately with interested users and providing short, intuitive explanations. An interactive question-answering process based on query refinement with semantic and contextual analysis enables the answering of more complex queries in natural language. Automatic tagging or classification reduces support costs when integrated into business processes.

Equation 3: Retrieval-Augmented Generation (RAG) probability decomposition

Let:

- x= user query
- d= retrieved context documents
- y= generated answer

Step 1: retrieval distribution

$$p(d \mid x)$$

Step 2: generation conditional on context

$$p(y \mid x, d)$$

Step 3: marginalize over retrieved contexts

$$p(y \mid x) = \sum_{d} p(y \mid x, d)\, p(d \mid x)$$

3.2. Contextualization and intent understanding

The knowledge of enterprise IT operations teams is often highly contextualized, with nuanced distinctions, short-lived relevance, and significant variability across teams, functions, industries, and companies. Consequently, even when a user query retrieves knowledge assets, those assets might not contain the required information. This mismatch may stem from missing context, as the user intention was to seek contextually similar assets but these were difficult to determine, or broken contextual links, where knowledge related to other knowledge is not connected and easily accessible, or from simply insufficient contextualization. Intention understanding is a deeper issue relating to the fact that even contextualizing queries properly, the full intention might still not be satisfied by a retrieved knowledge asset or a group of retrieved assets, and alternatives need to be suggested or additional knowledge discovered and shared. For example, one or multiple sources can be suggested for gathering further information; such sources can include the knowledge management system itself, other knowledge workers' profiles in the system, or external sources.

Academic research and corporate investments are increasingly focused on off-the-shelf products that apply the above techniques, especially for the content-free textual aspect. These products detect and generate text and messages in various alternative ways, such as bullets, summaries, or 2–3 levels of detail; they also convert natural-language input text into enterprise-standard-operational-business context with predefined or user-profile-specific vocabulary and abbreviations; and they include classification through tagging or customer-defined properties, automatic translations into and from the company's major spoken languages, emotional- and friendly-level conversion, and text-spelling correction. Such industrial AI is also more available for the audible aspect: multiple systems can now produce deepfake-like voices and synthesize individual speakers without expensive setups.

4. Architectural Considerations for AI-Driven KM Systems

The design of AI-driven knowledge management systems for enterprise IT operations requires consideration of architectural patterns that support effective use with AI, data governance practices to ensure knowledge consistency and quality, and the establishment of a suitable information model. Both knowledge retrieval and the expansion of user intent typically involve different models, with retrieval operating on document embeddings and intent expansion using a combination of domain knowledge, natural language processing, and machine learning techniques.

AI-Driven KM systems may be seen as integrated cyber-physical systems that combine a software layer with a corresponding collection of process and people components, all designed to enhance decision-making in enterprise IT operations. In general, such systems may be deployed using a dedicated service with real-time batch processing, using recent events to populate the knowledge base or using an AI engine that builds the knowledge periodically or in development mode. Knowledge management systems embedded into service desk tools may be better served with real-time processing, while operations dashboards typically require a dedicated service layer for real-time reporting on current issues whilst batched processing is used to indicate recurrent trends.

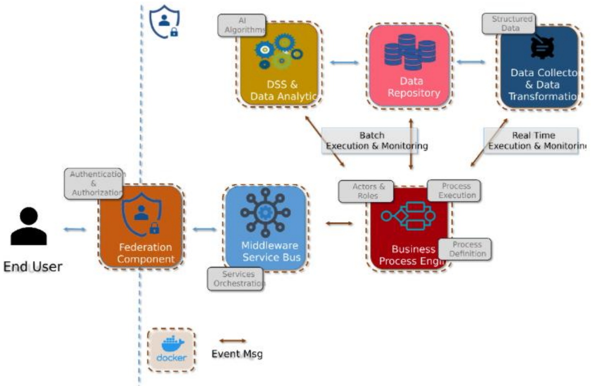

Fig 4: A Knowledge-Driven Framework for AI-Augmented

4.1. System architecture patterns

Artificial intelligence has propelled a considerable interest in knowledge management systems, especially in deployed in Integrated IT Operations platforms used to support Enterprise IT Operations teams. All existing implementations use AI technologies to enhance their information retrieval capabilities.

Depending on the use case, the specific architectural design will differ. The primary patterns have been extracted from the analysis of existing commercial products implementing AI-enabled knowledge management capabilities. For the AI-based systems aimed at information retrieval, the classic pattern of an inferencing engine search tool backed by a knowledge base in a semantic technology format has been followed. The knowledge base contains factual and temporal information which can be queried for simple document-like answers. The core innovation is the presence of the inferencing engine that helps improve the quality of the searches by complementing the search training data done during the preprocessing step. Text-to-Text transformation in the retrieval step is enabled by the use of Large Language Models (LLMs). The LLMs help in generating question-answer pairs from one-single passage documents, tag documents, and evaluate retrieved documents for their quality.

Systems that work in real-time mode and do not primarily focus on information retrieval exploit the specific characteristics of the knowledge management layer. The knowledge layer is used to store the past and present states of the Integrated IT Operations system. Such a layered architecture allows summers to discover the state of the system and the corresponding recommendation at any time. The present time information can be linked to the past information (e.g., root cause analysis) to extract new wrinkles. The knowledge base stores information retrieved directly from communications happening in the world and are documented by analysts and engineers operating in collaboration with the system.

Equation 4: Intent classification (NLP “intent understanding”)
Let features from query be $\mathbf{x}$. For multi-class intents $c \in \{1, \dots, C\}$:
Step 1: logits

$$z_c = \mathbf{w}_c^\top \mathbf{x} + b_c$$

Step 2: softmax to probabilities

$$p(c \mid \mathbf{x}) = \frac{e^{z_c}}{\sum_{r=1}^{C} e^{z_r}}$$

Step 3: predicted intent

$$\hat{c} = \arg\max_{c} p(c \mid \mathbf{x})$$

4.2. Data governance and quality

AI-Knowledge Management systems for enterprise IT operations ingested diverse sources without strict governance. KMS were updated on a schedule, often without a defined frequency. These shortcomings highlighted the significance of optimal data governance and quality control.

Improved data governance requires that the completeness, accuracy, relevance, and currency of incoming data be assured. To this end, it becomes essential to define data sources and models, frequency of ingestion, and processes to identify and rectify data quality issues. Specific roles can be assigned accountability for ensuring data quality. Even with advanced Large Language Models, absence of data quality control will often lead to incorrect conclusions and decisions.

Data quality is of paramount importance especially in real-time KMS serving frontline personnel. Delivery of the Service Desk constitutes an example. Responding to Service Desk tickets generally involves creating standard replies detailing reason and resolution of the underlying incidents. These responses over time are collated in a KMS and are used by the service agents for quick response. Service Desk responding to tickets related to account related issues is a high incidence and time-consuming demand on an IT organization extended during festival seasons when new users are onboarded. The accuracy of reply is sacrosanct as any breach in customer account credentials can lead to huge financial deficits. For such KM requirements, apart from standard ingestion from KMS, the KMS should enable real-time filtering and addition of relevant responses by the service agent before the reply is sent to the customer.

5. Deployment Models and Operational Implications

Deployment models for AI-driven knowledge management solutions vary along several dimensions, including on-premises versus cloud-based systems, hybrid configurations, and real-time versus batched processing of knowledge repositories. These choices not only determine the properties of the underlying hardware and software platforms but also have important implications for the day-to-day operations of enterprise IT functions.

Deploying AI-driven knowledge management solutions in a cloud infrastructure provides cost and management benefits, especially for smaller organizations, since the cloud service provider takes care of the underlying technology stack. Relying on a public cloud reduces capital expenditures, administration overhead, system maintenance, and upgrade efforts. Nevertheless, such services may also introduce concerns over data ownership and confidentiality. Use of hybrid clouds with both public and private portions can help to alleviate these concerns. Large enterprises with stringent requirements for the confidentiality of their knowledge repositories may, however, prefer to run the entire solution on their own infrastructure.

Equation 5: MTTR (Mean Time to Resolve) and MTTR reduction

If incident $i$has open time $t_{open}^{(i)}$and resolved time $t_{res}^{(i)}$:

Step 1: per-incident resolution time

$$\Delta t_i = t_{res}^{(i)} - t_{open}^{(i)}$$

Step 2: MTTR over Nincidents

$$MTTR = \frac{1}{N}\sum_{i=1}^{N} \Delta t_i$$

Step 3: Improvement from AI-KMS

$$\%\Delta MTTR = \frac{MTTR_{baseline} - MTTR_{AI}}{MTTR_{baseline}} \times 100$$

5.1. On-premises versus cloud-based deployments

Organizations can deploy an AI-based Knowledge Management (KM) system either on-premises on private or hybrid clouds, or through public cloud-based services. These two deployment models have different implications on the operational processes of IT operations teams. Using a public AI service like Open AI’s GPT or Azure OpenAI Service involves occasional data feeding to the service whenever the training needs to be updated. Private and hybrid-cloud-based service models still mandate the considerations for data governance and servicing the need for quality and secured data. In contrast, a fully on-premises AI KM environment necessitates having a Knowledge Base into any format that can be sent for training to either a generative AI service provider (for private cloud) or the LLMs which are available in open-source or community versions (for local installation).

On-premises AI KMS support processing the knowledge heatmap in real time; when an IT Operations Kick-off is raised, the team receives an auto-generated expertise matrix along with contextualized knowledge and KB info about the IT services involved in the request and the provider teams. All signals related to IT operations management (like incidents, SEs, and alerts) can be updated periodically or in real time. For the cloud or hybrid-cloud-base designs, it may not be possible to service the knowledge heatmap in real time, but a near real-time implementation is highly recommended for it being a critical service enhancer.

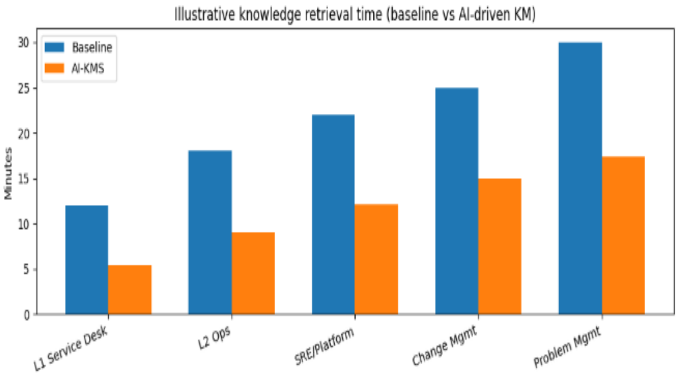

Fig 5: Comparison of Knowledge Retrieval Time: Baseline vs. AI-Driven KM

5.2. Real-time versus batched processing

Processing the data sources in real-time instead of on a batched schedule presents two principal benefits. First, it enables each operation of the IT function to have the most recent and relevant content at its disposal for any knowledge work, thus improving the overall quality and speed of this knowledge work. That, in turn, yields more effective conversation outcomes in the natural language interface to the KM system and ultimately drives the improvement of both customer and employee experience with the cloud-based services provided by the IT function. Second, it enables more timely responses to queries that present an unprecedented context. An organization regularly experiencing major public internet outages (natural disasters, etc.) may benefit from ingesting data on such events as they occur. Applying natural language models to generate conversational responses with that context will help ensure AML responses to customers are both correct and concise, thereby contributing both speed and quality dimensions to the customer experience delivery component of the NPS.

Many organizations have invested in implementing an SRE model that uses monitoring and alerting mechanisms to proactively identify issues. Consider the broader portfolio of services being operated by the IT function and the availability of data generated by those services. For example, an email service. Problems have been well-publicized in traditional media sources and social media channels. When availability or performance problems arise in email service, it is typical for internal systems and flagged queries to include information about service problems during that timeframe. Hence, why was automation of a public exchange and query with the new service information not achieved? Reducing the time lag from days or weeks in applying context data to serve customers to seconds can make a significant difference in customer experience.

6. Impact on IT Operations Teams

The transformation of enterprise IT environments through AI-driven knowledge management systems will have significant implications for IT operations teams. Senior management and organizational behavior research emphasizes the importance of knowledge work redesign when organizations deploy new digital technologies. The fundamental nature of the work undertaken by IT operations teams is changing: generalized, labor-intensive, operator-style work is declining while more complex decision management and testing work is increasing. As these teams deploy and operate increasingly integrated enterprise IT stacks, including cloud-based services and SaaS business applications, their member roles are evolving from a technology-oriented focus toward business process expertise. Many organizations are focusing on the business process areas that include digital capabilities, departing from a traditional, technology-focused structure in their IT operations teams. Collaboration across these areas is enabled and supported by tools that enhance knowledge sharing capabilities.

Real-time, just-in-time access to knowledge for these roles will improve the quality, consistency, and speed of delivery. AI-driven KM systems reduce the burden of contextualization and intent understanding in information retrieval processes, enabling team members to devote increased energy to the higher-value judgment and testing aspects of their responsibilities. However, AI still requires the specialized knowledge and experience of team members to evaluate and maintain source data, concerning both content quality and organizational relevance. To support these activities, team members require upskilling and reskilling in Knowledge Engineering, enabling them to develop a mix of tacit and explicit knowledge that describe and annotate the organization's knowledge base. While organizations can experiment with Knowledge Engineering practices — balancing skill-building and specific job role assignments AI-driven KM systems will increase the overall demand for team members skilled in this area.

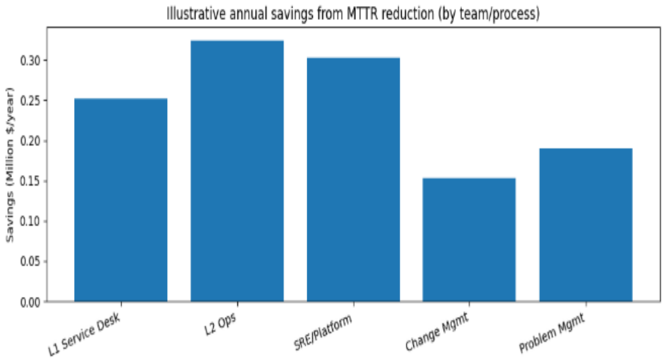

Fig 6: Annual Cost Savings from MTTR Reduction by Team

6.1. Knowledge work redesign and collaboration

AI-driven KM systems are poised to transform the design and practices associated with knowledge work in IT operations teams, improving task effectiveness and efficiency while also supporting collaboration. Knowledge work encompasses any work effort that, at least in part, requires a skillset that would not be possessed by the average worker and involves producing knowledge assets, such as how-to articles, rules for processing items, models, plans, and drawings (Drucker 1998). While enterprises use a broad set of enterprise tools and templates for measuring and managing quality, thus ensuring that one-time knowledge work assets are optimized for reuse in ongoing transactional work following consistent quality parameters, knowledge work within enterprise IT naturally makes use of heavy assets such as logical and physical data models, data dictionaries and data catalogs, and application portfolios. Thus, either transferring the assets to a knowledge management system (KMS) when no longer needed or associating with the shared marketplace for knowledge work in the organization needs to be viewed as the primary means of knowledge capture to optimize organization's capital and labor investments.

Some organizations develop a dedicated resource for overseeing, assessing, and validating the quality of, melding, and consolidating comparable knowledge work asset contributions into a coherent and receiveable form. While this investment typically generates higher-quality assets that are readily usable, proposed or alternative changes in the assets put forward for the knowledge marketplace continue to be evaluated. An AI approach that can detect the context and nature (sanity check) of a proposed change, and do so even on an ongoing and real-time basis during knowledge work execution, would service organizations that wish to steer a middle road of allowing for multiple contributions but blend-and-validate components into coherent forms.

6.2. Skill requirements and training

The impact of AI-driven knowledge management systems on IT operations teams goes beyond technology and processes. Knowledge work is redesigned through augmented intelligence systems that reduce the manual effort involved in knowledge writing and retrieval. Additional task automation offers opportunities to reduce short-term labor costs and reallocate technical knowledge workers to other activities. Their experience can be shared more widely and enable a different way of working, characterized by collaboration on knowledge verification or generation. The future skill requirements for IT operations teams are still emerging but certainly differ from traditional level-one or level-two support roles. The ITIL v4 guide, borne out of years of research and practice, anticipates the emerging skills needs associated with modern IT operations. The flow of information between AI systems and humans must be maintained, and workers need an overview of the current system state to detect problems promptly.

To realize these benefits, AI-driven knowledge management systems need to be deployed in production environments within IT organizations. Practical systems training and test support are essential for the successful deployeaent of AI systems in any domain. Most companies develop their IT operations-serving knowledge management systems as on-premises installations deployed either on local data center infrastructure or in dedicated segments of public cloud providers. In production, many knowledge management systems need to support real-time processing, because the knowledge in enterprise IT operations changes frequently and rapidly.

7. Conclusion

Enterprise IT operations, which focus on core enterprise applications, are becoming increasingly specialized and complex. These challenges have led many enterprises to apply AI-based knowledge management (KM) systems to address several key questions:

- How can IT operations teams manage knowledge work more efficiently?
- Can IT operations work be redefined to enable more efficient collaboration and knowledge sharing?
- What new skill sets are required of IT operations personnel, especially after deploying an AI-based KM system?

Owing in part to the diversity of deployment models and KM architectures, the impact of AI on knowledge management is likely to reach beyond traditional IT help desk operations to other areas of enterprise IT operations. Even at more advanced levels of AI application, it is still beneficial, if not essential, to augment AI with basic KM principles, such as content moderation, active collaboration across domain silos, and user education. This enables enterprises to achieve more effective deployment of IT operations knowledge within their respective organizations.

Table 1: Illustrative ROI table

Team/Process	Annual Tickets	Blended Cost ($/min)	Annual Time Saved (min)
L1 Service Desk	5000	1.2	210000.0
L2 Ops	2500	1.8	180000.0
SRE/Platform	1200	2.5	121200.0
Change Mgmt	800	2.0	76800.0
Problem Mgmt	400	3.0	63200.0

7.1. Future Trends

Current trends in knowledge management for IT operations include the integration of AI technology to support the knowledge creation, storage, and sharing processes. Nevertheless, the formalization of knowledge management is not widespread; many organizations lack the tools or procedures required to effectively manage knowledge. In response, the increasing prevalence of cloud platforms with built-in capabilities for managing the creation and authentication of content has fostered the development of applications that leverage existing content to enhance the relevance of search results. AI technology is also being used to enhance knowledge management by adding a contextualization layer to information retrieval, thereby improving the understanding of user intent and the relevance of results.

There is considerable interest in addressing the knowledge gap within IT operations teams due to insufficient documentation, challenges with information retrieval tools, and lack of context and maintenance of existing logs. AI-driven solutions for managing knowledge in service desks, incident management, and observability systems are emerging to enhance search results and enable the auto-generation of knowledge articles. These solutions overcome the existing problems of previous generations. Nevertheless, the current focus on enhancing information retrieval processes has attracted most attention, while research and deployed systems capable of covering the entire knowledge lifecycle remain scarce. Looking ahead, systems that automatically manage knowledge generation, storage, sharing, and usage will soon be ready for deployment. The first wave of such systems is expected to emerge from service desks and incident management solutions.

References

[1] Agrawal, R., & Srikant, R. (2000). Privacy-preserving data mining. ACM SIGMOD Record, 29(2), 439–450.

[2] Meda, R. (2023). Developing AI-Powered Virtual Color Consultation Tools for Retail and Professional Customers. Journal for ReAttach Therapy and Developmental Diversities. https://doi. org/10.53555/jrtdd. v6i10s (2), 3577.

[3] Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19–31.

[4] Gottimukkala, V. R. R. (2023). Privacy-Preserving Machine Learning Models for Transaction Monitoring in Global Banking Networks. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 633-652.

[5] Aljawarneh, S., Aldwairi, M., & Yassein, M. B. (2018). Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. Journal of Computational Science, 25, 152–160.

[6] Kummari, D. N. (2023). Energy Consumption Optimization in Smart Factories Using AI-Based Analytics: Evidence from Automotive Plants. Journal for Reattach Therapy and Development Diversities. https://doi. org/10.53555/jrtdd. v6i10s (2), 3572.

[7] Bates, D. W., Saria, S., Ohno-Machado, L., et al. (2014). Big data in health care. Health Affairs, 33(7), 1123–1131.

[8] Keerthi Amistapuram. (2023). Privacy-Preserving Machine Learning Models for Sensitive Customer Data in Insurance Systems. Educational Administration: Theory and Practice, 29(4), 5950–5958. https://doi.org/10.53555/kuey.v29i4.10965

[9] Belle, A., Thiagarajan, R., Soroushmehr, S. M. R., et al. (2015). Big data analytics in healthcare. BioMed Research International, 2015, 370194.

[10] Guntupalli, R. (2023). AI-Driven Threat Detection and Mitigation in Cloud Infrastructure: Enhancing Security through Machine Learning and Anomaly Detection. Available at SSRN 5329158.

[11] Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. ACM SIGMOD Record, 29(2), 93–104.

[12] Unifying Data Engineering and Machine Learning Pipelines: An Enterprise Roadmap to Automated Model Deployment. (2023). American Online Journal of Science and Engineering (AOJSE) (ISSN: 3067-1140) , 1(1). https://aojse.com/index.php/aojse/article/view/19.

[13] Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile Networks and Applications, 19(2), 171–209.

[14] Siva Hemanth Kolla. (2023). Deep Learning–Driven Retrieval-Augmented Generation for Enterprise ITSM Automation: A Governance-Aligned Large Language Model Architecture. Journal of Computational Analysis and Applications (JoCAAA), 31(4), 2489–2502. Retrieved from https://www.eudoxuspress.com/index.php/pub/article/view/4774.
[15] Cios, K. J., & Moore, G. W. (2002). Uniqueness of medical data mining. Artificial Intelligence in Medicine, 26(1–2), 1–24.
[16] Kummari, D. N., & Burugulla, J. K. R. (2023). Decision Support Systems for Government Auditing: The Role of AI in Ensuring Transparency and Compliance. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 493-532.
[17] Dasgupta, D., & Nino, F. (2009). Immunological computation. CRC Press.
[18] Varri, D. B. S. (2023). Advanced Threat Intelligence Modeling for Proactive Cyber Defense Systems. Available at SSRN 5774926.
[19] Dwork, C. (2008). Differential privacy. ICALP Proceedings, 1–12.
[20] El Emam, K., & Dankar, F. K. (2008). Protecting privacy using k-anonymity. JAMIA, 15(5), 627–637.
[21] Meda, R. (2023). Data Engineering Architectures for Scalable AI in Paint Manufacturing Operations. European Data Science Journal (EDSJ) p-ISSN 3050-9572 en e-ISSN 3050-9580, 1(1).
[22] Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874.
[23] Friedman, C., & Elhadad, N. (2014). Natural language processing in health care. In Biomedical Informatics. Springer.
[24] Garapati, R. S. (2022). AI-Augmented Virtual Health Assistant: A Web-Based Solution for Personalized Medication Management and Patient Engagement. Available at SSRN 5639650.
[25] Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms. Pattern Recognition, 64, 206–223.
[26] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
[27] Segireddy, A. R. (2021). Containerization and Microservices in Payment Systems: A Study of Kubernetes and Docker in Financial Applications. Universal Journal of Business and Management, 1(1), 1–17. Retrieved from https://www.scipublications.com/journal/index.php/ujbm/article/view/1352
[28] He, J., Baxter, S. L., Xu, J., et al. (2019). The practical implementation of AI in healthcare. Nature Medicine, 25(1), 30–36.
[29] Inala, R. AI-Powered Investment Decision Support Systems: Building Smart Data Products with Embedded Governance Controls.
[30] Hripcsak, G., & Albers, D. J. (2013). Next-generation phenotyping. JAMIA, 20(1), 117–121.
[31] Gottimukkala, V. R. R. (2021). Digital Signal Processing Challenges in Financial Messaging Systems: Case Studies in High-Volume SWIFT Flows.
[32] Iglewicz, B., & Hoaglin, D. C. (1993). How to detect and handle outliers. ASQC.
[33] Johnson, A. E. W., Pollard, T. J., Shen, L., et al. (2016). MIMIC-III database. Scientific Data, 3, 160035.
[34] Yandamuri, U. S. (2022). Big Data Pipelines for Cross-Domain Decision Support: A Cloud-Centric Approach. International Journal of Scientific Research and Modern Technology, 1(12), 227–237. https://doi.org/10.38124/ijsrmt.v1i12.1111
[35] Kimball, R., & Caserta, J. (2004). The data warehouse ETL toolkit. Wiley.
[36] Davuluri, P. N. Integrating Artificial Intelligence into Event-Driven Financial Crime Compliance Platforms.
[37] Kriegel, H. P., Kröger, P., Schubert, E., & Zimek, A. (2009). Outlier detection in axis-parallel subspaces. PKDD Proceedings, 831–838.
[38] Kummari, D. N. (2023). AI-Powered Demand Forecasting for Automotive Components: A Multi-Supplier Data Fusion Approach. European Advanced Journal for Emerging Technologies (EAJET)-p-ISSN 3050-9734 en e-ISSN 3050-9742, 1(1).
[39] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.
[40] Li, Y., Chen, C. Y., Wasserman, W. W., & Ramani, A. K. (2016). Deep feature selection. Bioinformatics, 32(5), 743–750.
[41] Goutham Kumar Sheelam, Hara Krishna Reddy Koppolu. (2022). Data Engineering And Analytics For 5G-Driven Customer Experience In Telecom, Media, And Healthcare. Migration Letters, 19(S2), 1920–1944. Retrieved from https://migrationletters.com/index.php/ml/article/view/11938
[42] Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long short-term memory networks for anomaly detection. ESANN Proceedings.
[43] Mandl, K. D., & Kohane, I. S. (2015). Data sharing in healthcare. BMJ, 350, h988.
[44] Garapati, R. S. (2023). Optimizing Energy Consumption in Smart Build-ings Through Web-Integrated AI and Cloud-Driven Control Systems.
[45] Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare. Briefings in Bioinformatics, 19(6), 1236–1246.
[46] Kushvanth Chowdary Nagabhyru. (2023). Accelerating Digital Transformation with AI Driven Data Engineering: Industry Case Studies from Cloud and IoT Domains. Educational Administration: Theory and Practice, 29(4), 5898–5910. https://doi.org/10.53555/kuey.v29i4.10932
[47] Murphy, S. N., Weber, G., Mendis, M., et al. (2010). i2b2 platform. JAMIA, 17(2), 124–130.

[48] Ramesh Inala. (2023). Big Data Architectures for Modernizing Customer Master Systems in Group Insurance and Retirement Planning. Educational Administration: Theory and Practice, 29(4), 5493–5505. https://doi.org/10.53555/kuey.v29i4.10424

[49] Patcha, A., & Park, J. M. (2007). An overview of anomaly detection techniques. Computer Networks, 51(12), 3448–3470.

[50] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn. Journal of Machine Learning Research, 12, 2825–2830.

[51] Aitha, A. R. (2023). CloudBased Microservices Architecture for Seamless Insurance Policy Administration. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 607-632.

[52] Rajkomar, A., Oren, E., Chen, K., et al. (2018). Scalable deep learning with EHRs. NPJ Digital Medicine, 1, 18.

[53] Avinash Reddy Segireddy. (2022). Terraform and Ansible in Building Resilient Cloud-Native Payment Architectures. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 444–455. Retrieved from https://www.ijisae.org/index.php/IJISAE/article/view/7905.

[54] Ringberg, H., Soule, A., Rexford, J., & Diot, C. (2007). Sensitivity of PCA for anomaly detection. SIGMETRICS Proceedings.

[55] Koppolu, H. K. R., Sheelam, G. K., & Komaragiri, V. B. (2023). Autonomous Telecommunication Networks: The Convergence of Agentic AI and AI-Optimized Hardware. International Journal of Science and Research (IJSR), 12(12), 2253-2270.

[56] Ruff, L., Vandermeulen, R. A., Görnitz, N., et al. (2018). Deep one-class classification. ICML Proceedings.

[57] Rongali, S. K. (2023). Explainable Artificial Intelligence (XAI) Framework for Transparent Clinical Decision Support Systems. International Journal of Medical Toxicology and Legal Medicine, 26(3), 22-31.

[58] Salfner, F., Lenk, M., & Malek, M. (2010). Survey of failure prediction methods. ACM Computing Surveys, 42(3), 1–42.

[59] Nagubandi, A. R. (2023). Advanced Multi-Agent AI Systems for Autonomous Reconciliation Across Enterprise Multi-Counterparty Derivatives, Collateral, and Accounting Platforms. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 653-674.

[60] Schölkopf, B., Platt, J. C., Shawe-Taylor, J., et al. (2001). Estimating the support of a high-dimensional distribution. Neural Computation, 13(7), 1443–1471.

[61] Uday Surendra Yandamuri. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. International Journal Of Finance, 36(6), 682-706. https://doi.org/10.5281/zenodo.18095256

[62] Sipos, R., Fradkin, D., Moerchen, F., & Wang, Z. (2014). Log-based predictive maintenance. KDD Proceedings.

[63] Meda, R. (2023). Intelligent Infrastructure for Real-Time Inventory and Logistics in Retail Supply Chains. Educational Administration: Theory and Practice.

[64] Kolla, S. K. (2021). Designing Scalable Healthcare Data Pipelines for Multi-Hospital Networks. World Journal of Clinical Medicine Research, 1(1), 1–14. Retrieved from https://www.scipublications.com/journal/index.php/wjcmr/article/view/1376

[65] Bandi, V. D. V. K. (2023). Cloud-Native Model Lifecycle Management for Enterprise AI Systems. International Journal of Scientific Research and Modern Technology, 2(12), 78–90. https://doi.org/10.38124/ijsrmt.v2i12.1236

[66] Inala, R. Revolutionizing Customer Master Data in Insurance Technology Platforms: An AI and MDM Architecture Perspective.

[67] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society B, 58(1), 267–288.

[68] Garapati, R. S. (2022). Web-Centric Cloud Framework for Real-Time Monitoring and Risk Prediction in Clinical Trials Using Machine Learning. Current Research in Public Health, 2, 1346.

[69] Tukey, J. W. (1977). Exploratory data analysis. Addison-Wesley.

[70] AI Powered Fraud Detection Systems: Enhancing Risk Assessment in the Insurance Sector. (2023). American Journal of Analytics and Artificial Intelligence (ajaai) With ISSN 3067-283X, 1(1). https://ajaai.com/index.php/ajaai/article/view/14

[71] Weber, G. M., Mandl, K. D., & Kohane, I. S. (2014). Finding the missing link for big biomedical data. JAMIA, 21(1), 1–3.

[72] Kolla, S. H. (2021). Rule-Based Automation for IT Service Management Workflows. Online Journal of Engineering Sciences, 1(1), 1–14. Retrieved from https://www.scipublications.com/journal/index.php/ojes/article/view/1360

[73] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). FAIR Guiding Principles. Scientific Data, 3, 160018.

[74] Zhang, Y., & Yang, Q. (2021). A survey on multi-task learning. IEEE Transactions on Knowledge and Data Engineering, 34(12), 5586–5609.

[75] Gottimukkala, V. R. R. (2022). Licensing Innovation in the Financial Messaging Ecosystem: Business Models and Global Compliance Impact. International Journal of Scientific Research and Modern Technology, 1(12), 177-186.

[76] Zhou, Z. H. (2012). Ensemble methods. CRC Press.

[77] Guntupalli, R. (2023). Optimizing Cloud Infrastructure Performance Using AI: Intelligent Resource Allocation and Predictive Maintenance. Available at SSRN 5329154.

[78] Little, R. J. A., & Rubin, D. B. (2002). Statistical analysis with missing data. Wiley.

[79] Siva Hemanth Kolla. (2022). Knowledge Retrieval Systems for Enterprise Service Environments. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 495–506. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/8037

[80] Bishop, C. M. (1994). Novelty detection and neural network validation. IEE Proceedings, 141(4), 217–222.

[81] Rongali, S. K. (2022). AI-Driven Automation in Healthcare Claims and EHR Processing Using MuleSoft and Machine Learning Pipelines. Available at SSRN 5763022.

[82] Cook, D. J., & Holder, L. B. (2006). Mining graph data. Wiley.

[83] Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). Time series analysis: Forecasting and control. Wiley.

[84] Amistapuram, K. (2022). Fraud Detection and Risk Modeling in Insurance: Early Adoption of Machine Learning in Claims Processing. Available at SSRN 5741982.

[85] Hyndman, R. J. (2020). Forecasting principles. Journal of Statistical Software, 27(3), 1–22.

[86] Varri, D. B. S. (2022). A Framework for Cloud-Integrated Database Hardening in Hybrid AWS-Azure Environments: Security Posture Automation Through Wiz-Driven Insights. International Journal of Scientific Research and Modern Technology, 1(12), 216-226.

[87] Aggarwal, C. C. (2017). Outlier analysis (2nd ed.). Springer.

[88] Davuluri, P. N. AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems.

[89] Bifet, A., & Gavalda, R. (2007). Learning from time-changing data with adaptive windowing. SDM Proceedings.

[90] Nagabhyru, K. C. (2023). From Data Silos to Knowledge Graphs: Architecting CrossEnterprise AI Solutions for Scalability and Trust. Available at SSRN 5697663.

[91] Zaharia, M., Chowdhury, M., Franklin, M. J., et al. (2010). Spark: Cluster computing. HotCloud Proceedings.

[92] Avinash Reddy Aitha. (2022). Deep Neural Networks for Property Risk Prediction Leveraging Aerial and Satellite Imaging. International Journal of Communication Networks and Information Security (IJCNIS), 14(3), 1308–1318. Retrieved from https://www.ijcnis.org/index.php/ijcnis/article/view/8609

[93] Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing. Communications of the ACM, 51(1), 107–113.

[94] Kolla, S. K. (2021). Architectural Frameworks for Large-Scale Electronic Health Record Data Platforms. Current Research in Public Health, 1(1), 1–19. Retrieved from https://www.scipublications.com/journal/index.php/crph/article/view/1372

[95] Bandi, V. D. V. K. (2023). Production-Grade Machine Learning Pipelines For Healthcare Predictive Analytics. South Eastern European Journal of Public Health, 189–205. Retrieved from https://www.seejph.com/index.php/seejph/article/view/7057