How Generative AI Impacts Knowledge Management

Published 29 November 2023 - ID G00798807 - 15 min read

By Analyst(s): Haritha Khandabattu, Stephen Emmott, Darin Stewart

Initiatives: Artificial Intelligence

Capturing, sharing and retrieving knowledge is central to knowledge management and essential to employee effectiveness, yet is rarely done well. Application leaders providing technology that supports knowledge management must begin to leverage GenAI to viably automate and augment key KM activities.

Overview

Impacts

- Generative AI's (GenAI's) ability to turn byproducts of normal interactions into structured knowledge assets radically reduces the effort required to capture institutional and subject matter knowledge in external, persistent forms.
- Harnessing large language models significantly improves knowledge workers' productivity by reducing the significant amount of their time that they spend searching for and making sense of information.
- Generative AI's capability to automatically extract data from documents and messages to produce new content promises resolution of a long-standing obstacle in knowledge management (KM): building and maintaining expertise directories that help employees find and connect with one another for knowledge sharing and collaboration.

Recommendations

- Use large language models (LLMs) and generative AI services to summarize interaction transcripts, extract key points and format the result in accordance with enterprise style guidelines and templates.
- Transform your enterprise search by deploying LLMs, using prompt engineering techniques like in-context learning to enable improved search experience and improve search accuracy.
- Shift to a dynamic expertise directory by utilizing generative AI services to continuously profile expertise from the content of documents, messages and other sources.

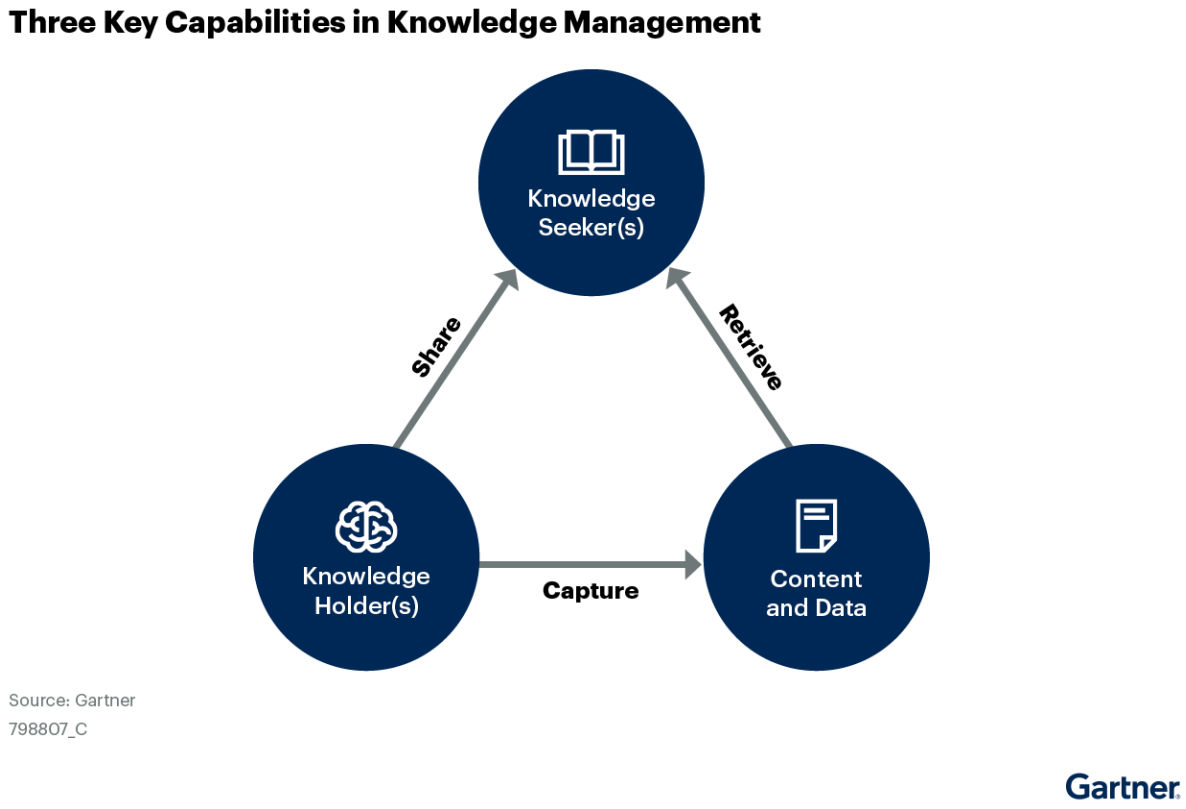
Introduction

Generative AI is one of the most disruptive technology trends to hit the global market in decades. With its advancements, Knowledge management is again at the forefront of application leaders' minds. Gartner research shows that 79% of leaders believe knowledge management and insight are extremely or very important to achieving their organization's goals for 2023. [1]

Although application leaders are ready to learn and reap the benefits of GenAI in knowledge management, they have little to no guidance on where to get started. This note provides that guidance by focusing on three key capabilities of knowledge management — Capture, Retrieve and Share. The note explains how GenAI can elevate each capability, enabling leaders to extract maximum value from their knowledge assets and drive informed decision making.

Figure 1 shows three key capabilities in knowledge management that are impacted by Generative AI.

Figure 1: Three Key Capabilities in Knowledge Management



Source: Gartner
798807_C

Gartner

Table 1: How Generative AI Impacts Knowledge Management

(Enlarged table in Appendix)

Impacts	Top Recommendations
Generative AI's ability to turn byproducts of normal interactions into structured knowledge assets radically reduces the effort required to capture institutional and subject matter knowledge in external, persistent forms.	■ Use large language models and Generative AI services to summarize interaction transcripts, extract key points and format the result in accordance with enterprise style guidelines and templates.
Harnessing the large language models significantly improves the productivity of knowledge workers who usually spend a significant amount of their time per day searching for information and making sense of it.	■ Transform your enterprise search by deploying LLMs to enable improved search experience and improve search accuracy.
Generative AI's capability to automatically extract data from documents and messages to produce new content promises resolution of a long-standing obstacle in knowledge management: building and maintaining expertise directories that help employees find and connect with one another for knowledge sharing and collaboration.	■ Shift to an expertise directory that uses generative AI to dynamically profile expertise by first reviewing your enterprise application portfolio, as well as what is available to buy, before developing your own solution atop an appropriate platform.

Source: Gartner (November 2023)

Impacts and Recommendations

LLMs and GenAI Can Facilitate Knowledge Capture

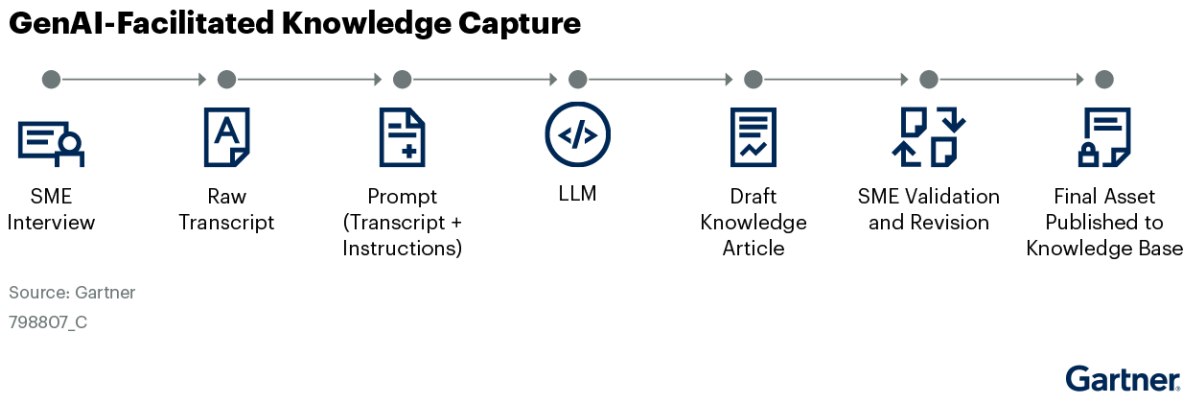
Large language models (LLMs) and GenAI are directly applicable to knowledge capture; that is, turning tacit knowledge — what's in someone's head — into explicit knowledge, captured in some external, persistent form like a knowledge base article. This is a fundamental aspect of knowledge management and remains KM's biggest challenge. Subject matter experts (SMEs) are busy and rarely have time to sit down and codify their hard-won expertise into a document or wiki entry. LLMs and GenAI can reduce the effort required to capture that knowledge and expertise.

Generative AI can use meeting and conversation transcripts as the basis for knowledge assets. One collateral benefit of the move to remote or hybrid work is that most interactions are now conducted over video collaboration tools like Teams, Webex or Zoom. Other collaboration tools, such as email, workstream collaboration and chat transcripts, are also potential source material. Each tool can produce a transcript of those interactions. Those transcripts are excellent raw material for GenAI summarization, extraction and formatting.

Generative AI can extract knowledge assets from transcripts of meetings and other interactions freeing experts of the need to explicitly write or capture their knowledge.

Rather than tasking a SME to write down everything they know about a topic, they can simply talk about it into a webcam. Alternatively, a colleague who would benefit from the knowledge could interview the SME. In addition to the benefit of immediate knowledge transfer from the SME to their colleague, that conversation can be summarized with key points extracted, and the result formatted to the knowledge base content standard (see Figure 2). In addition to such explicit knowledge capture efforts, transcripts of end-of-sprint demonstrations, product demonstrations and feature walk-through could be similarly captured.

Figure 2: GenAI-Facilitated Knowledge Capture

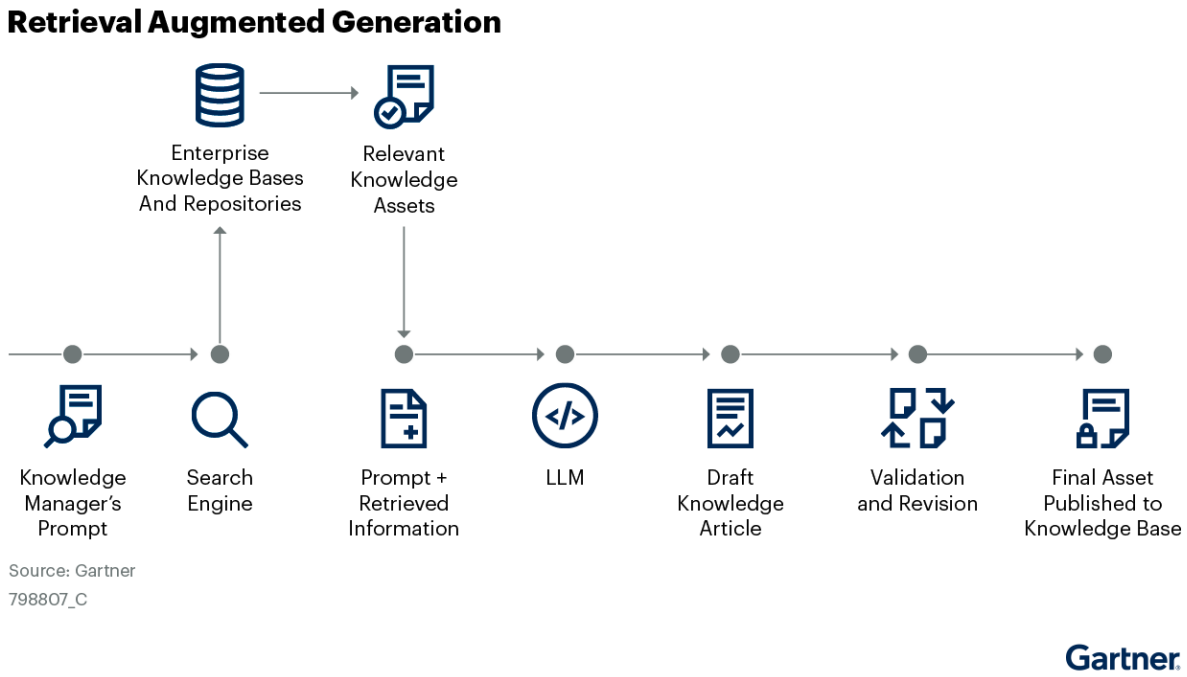


Knowledge capture can be taken a step further with existing explicit knowledge resources, such as knowledge bases, wiki articles or any other content repository. Retrieval Augmented Generation (RAG) uses a search engine to locate and retrieve enterprise resources to be passed to an LLM as part of a prompt, along with instructions describing what to do with that content. This allows knowledge resources from across the enterprise to be used as raw material for GenAI. Multiple knowledge resources pertaining to a particular topic or issue can be gathered, recombined and consolidated into a single asset or restructured in multiple forms. The processes and infrastructure supporting this sort of capability can be complex but can facilitate a wide range of use cases.

(For a full discussion of Retrieval Augmented Generation, see Prompt Engineering With Enterprise Information for LLMs and GenAI.)

As an example of retrieval augmented generation, closed help desk tickets with successful resolutions to specific issues could be retrieved and included in a prompt to an LLM, along with instructions to identify the common solution. The resulting content could then be used as the basis of a knowledge base article, self-service/self-solve content or any other desired presentation. The LLM can be instructed to derive multiple formats from the same source (see Figure 3). This process can also be used to keep knowledge assets current. As new knowledge resources become available, they can be retrieved and passed to the LLM, along with the original, consolidated content. The LLM could then update the resource to include the updated information.

Figure 3: Retrieval Augmented Generation



Any AI-generated content, even previously published information that is only being updated, needs to be reviewed and validated by a human being. An LLM will formulate an answer that sounds natural. It will not ensure the response is accurate. Additionally, how the prompt itself is constructed can have a dramatic impact on the content that is generated. Minor changes in wording or instructions in the prompt can cause completely different and often contradictory results from the LLM. Before any extracted or captured knowledge is published, it needs to be checked by a human — preferably the SME on the topic.

Once the knowledge asset is created, formatted and validated, it must be maintained and managed. These processes require consistent metadata to facilitate both administration and findability. Here again, the LLM can be called upon to select and apply appropriate metadata for the knowledge asset that was just created. Ideally, a standard set of metadata elements will be made available to LLM, along with term lists containing valid values for those metadata fields. Doing so will, if not ensure consistent metadata, at least improve and simplify metadata application, which will in turn improve the overall knowledge environment.

Recommendations:

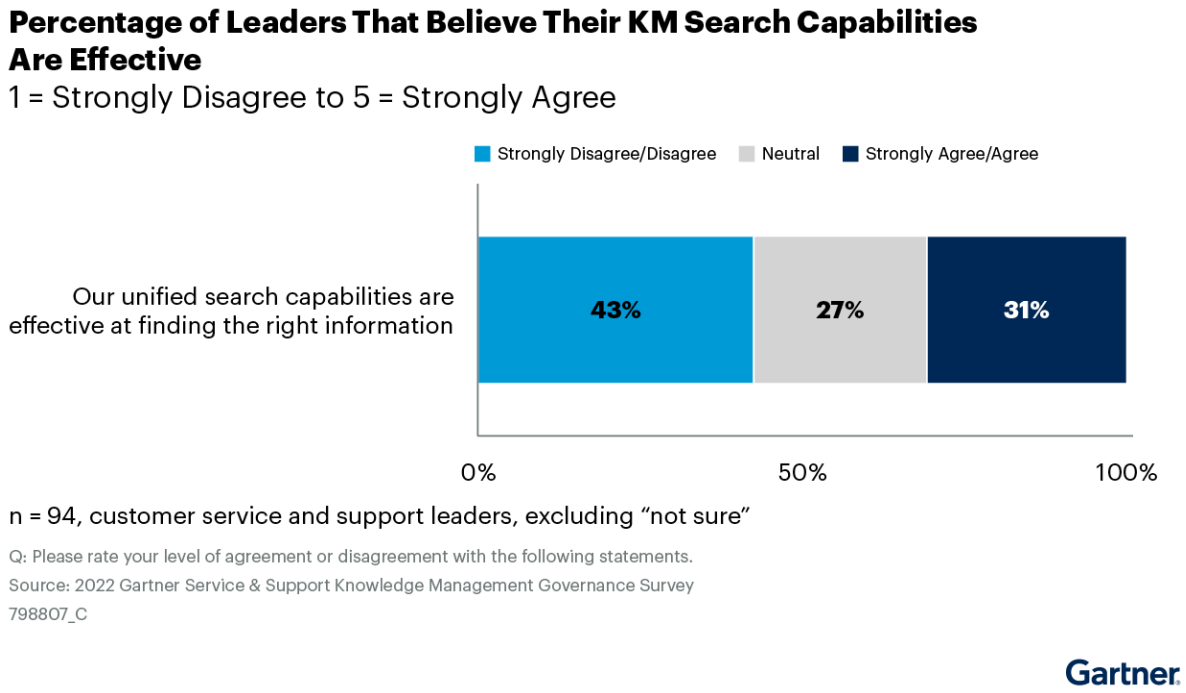
- Use large language models and Generative AI services to act as both scribe and summarizer, turning byproducts of normal interactions into structured knowledge assets. Automatically extract key points and format the result in accordance with enterprise-style guidelines and templates.

Large Language Models Enhance Search and Retrieval

Every enterprise has unique knowledge search requirements, distinct from those of general users on platforms like Google, where search methods follow a linear approach. In a linear approach, queries are matched against a dataset, with results ranked based on similarity. In an enterprise setting, this approach may *not* suffice to meet the unique search needs of employees seeking precise and relevant internal information. For instance: an employee seeking information in an enterprise setting must be able to search for information without compromising compliance with regulations like GDPR, HIPAA or company-specific data governance policies.

Enterprise search systems demand specificity in internal queries, eliminating redundant and vague information. These systems serve as components of dialogue, question-answering, and recommender systems. In Gartner's 2022 Service and Support Knowledge Management Governance survey, fewer than a third of leaders believe that their unified search capabilities are effective at finding the right information (see Figure 4). [1]

Figure 4: Percentage of Leaders That Believe Their KM Search Capabilities Are Effective

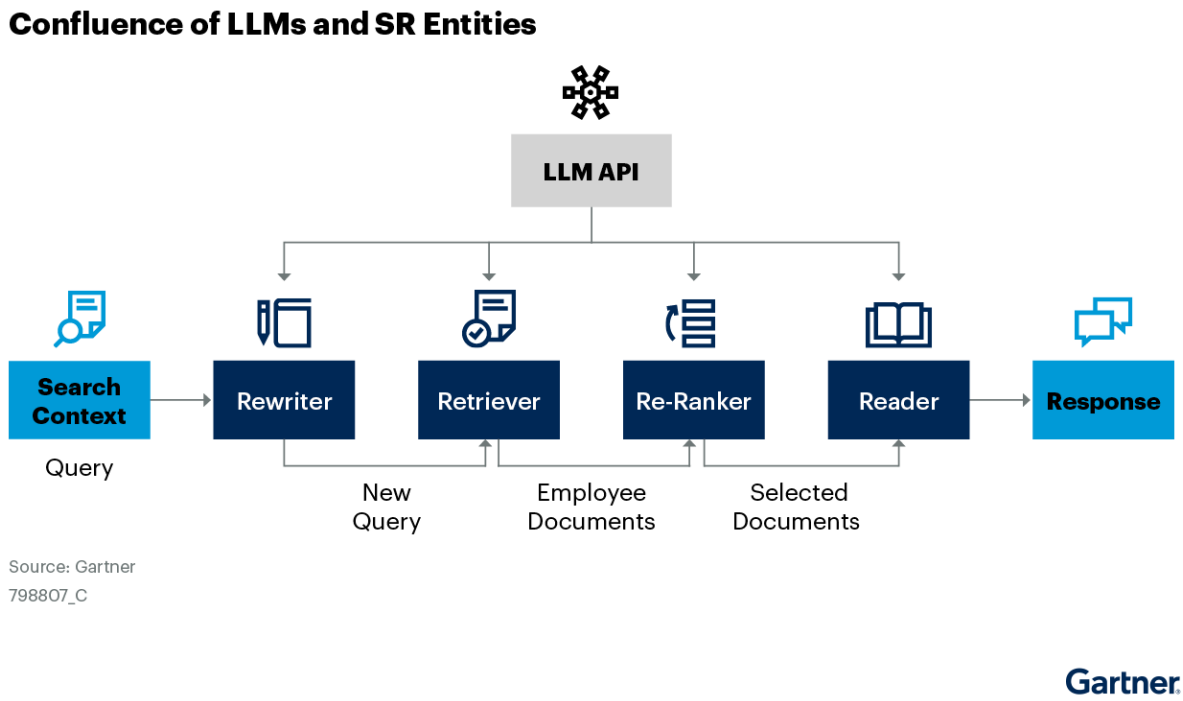


The trajectory of search and retrieval systems (SR) has evolved dramatically from its origins in term-based methods to its integration with advanced deep learning techniques like semantic vector search. [2] The deep learning techniques excel at capturing complex contextual signals and semantic nuances, thereby reshaping the SR landscape. The evolution requires a combination of traditional methods (i.e., term-based sparse retrieval methods with rapid response) and modern neural architectures (i.e., large language models with powerful language understanding capacity and entity relationship databases — the same as vector DBS).

One prominent method of applying LLMs within enterprise SR is through in-context learning (ICL). [3] There are four entities in the traditional SR system — query rewriters, retrievers, rerankers and readers.

Figure 5 shows the confluence of LLMs and SR entities.

Figure 5: Confluence of LLMs and SR Entities



- **Query Rewriter** module refines user queries for better precision and expression. Traditional rewriters modify queries to closely match user intent based on predefined rules. LLMs augment query writing capabilities by fully capturing the subtle nuances of user intent.
- **Retriever** acts as the first-pass document filter to collect broadly relevant documents for user queries. LLM will help the retriever module exhibit superior semantic capability and excel at recognizing complicated user intent.
- **Reranker** aims to rerank a document list retrieved by the retriever based on the query-document relevance. LLMs can help reranking by incorporating large datasets, providing explanations for the ranking and improving interpretability.
- **Reader** generates natural language responses based on the document corpus. By providing documents continuously to the LLMs, the generated responses are more accurate and information-rich than the original retrieved lists.

This integration of LLMs and deep learning techniques in the SR ecosystem represents a paradigm shift in natural language processing due to their remarkable language understanding, generation, generalization and reasoning abilities. However, semantic vector search sometimes has poor search quality — often because the knowledge documents are too dense and not representative of the LLMs' training data. To enhance search quality, you can explore keyword-based neural search methods, improving matches by identifying specific keywords for high-relevance content chunks. Additionally, enriching metadata with contextual information about the chunk's surroundings can further improve results.

LLMs can also be used in a variety of ways to improve fault tolerance of SR systems. First, they can detect and diagnose faults more accurately than traditional methods. Second, they can recover from faults more quickly and efficiently. Finally, they can proactively prevent faults from occurring in the first place. In short, LLMs can help improve reliability and availability of SR systems.

Recommendations:

- Transform your enterprise search by deploying LLMs to enable improved search experience and improve search accuracy.

Finding People to Share Knowledge Becomes Achievable

An organization's principal repositories of knowledge are the brains of its employees. Leveraging this knowledge, for the employees who seek it, remains a perennial challenge for all organizations. The focus of knowledge management initiatives — and that of the vendors providing tools — tends to be on the capture, storage and retrieval of knowledge made explicit in the form of content and data. Sharing knowledge directly or indirectly between people is often neglected, with few tools available to help employees find each other based on what and who they know. Sometimes, this is referred to as *expertise location* (see Quick Answer: Help Employees Find Each Other to Share Knowledge).

Various applications — both bought and built — across many markets provide capabilities that can meet varying levels of connection between employees. Such tools are centered around representing employees with a profile that captures who they are, what they know and who they know. However, in addition to few vendors or products providing these capabilities, the onus is upon the employer — sometimes employees — to create and maintain the profiles. The number of profiles, and the frequency of change, require either a dedicated and substantial team or the discretionary contribution of employees to create and maintain their profiles. Unsurprisingly, this is where best intentions meet reality. Such initiatives fail to scale in all, but the smallest or most determined organizations due to the manual work required, which is why Generative AI will have a substantial impact on this aspect of knowledge management. Initially, this will be in terms of build, but later a buy approach as relevant vendors accommodate generative AI capabilities into their products.

Despite being effectively black boxes frozen in terms of learning, the output (i.e., responses) of LLMs can be steered by engineering the prompts presented to them. This makes possible a range of natural language processing capabilities — including classification, extraction, generation and summarization. Precisely the capabilities needed to manually maintain expertise profiles.

Extraction of Facts From Connected Data

The content of expertise profiles can be sourced from the myriad files, messages and applications that users interact with. For example:

- Basic *who's who* information can be sourced from human capital management systems (via an org chart)
- *Who knows what* from the contents of files directly owned or shared in collaborative management systems
- *Who knows who* from collaborative work management systems detailing active projects.

Using an insight engine to appropriately crawl and index such data sources enables matching documents to be identified using vectorization. Once identified, digital assets appropriate to the data sought (e.g., current projects) can be identified and fed into a prompt with instructions for what to extract and the required format (e.g., a list of projects by lead and contributor, and how to present this). The extracted data can then be stored in a secondary data model for the generation of profiles. Generative AI enables tasks that would require complex programming by specialists to be achieved by generalists through natural language.

Creation of Profiles

While structured data can be used to generate profiles without generative AI by using structured templates, there are occasions where flexibility is needed. For instance, presenting profiles — in whole or part — as part of a conversation or embedded in the context of third-party applications. In these situations, LLMs can be prompted with secondary data sources to generate profiles or their parts. It may also be desirable to forgo the need for a secondary data source, and the extractions required to populate this. LLMs can take primary data from a variety of sources and generate profiles to specification in a single step. For instance, taking the user profiles from three applications — collaborative work management, workstream collaboration and human capital management system (e.g., org chart) — generates an expertise profile. Using primary data sources, such as Microsoft Graph, expands the possibilities even further.

Facilitating Retrieval With Metadata

Key to successful search and retrieval is good quality metadata (i.e., keywords that comprehensively and unambiguously present the content they are attributed to). LLMs provide a potential solution to the long-standing challenge of automatically attributing metadata to data sources — especially content-based sources, such as documents and videos. Given only the salient sections of a document, such as title, headings, first and last paragraphs as input, a suitably engineered prompt can result in structured metadata representing the document (e.g., title, author and keywords derived from a fixed list of subjects). These can be attributed to content and data assets or profiles themselves.

Generative AI's capability to automatically extract data from content to produce new content promises resolution of a long-standing obstacle in knowledge management: scalable expertise directories listing who's who, who knows what and who knows who. Finding and connecting with others is essential for knowledge sharing and collaboration, but few solutions exist at the time of writing.

Recommendations:

- Establish whether you have a solution already available by reviewing your enterprise application portfolio. Also, consider developing your own solution by developing the capabilities/application atop an appropriate platform.
- Determine whether an appropriate solution is available to buy by reviewing relevant markets (see Quick Answer: Help Employees Find Each Other to Share Knowledge) for more information.

Evidence

[1] **2022 Gartner Service and Support Knowledge Management Governance Survey.** This survey was conducted online from 8 November through 18 November to understand the factors most critical for effective knowledge management processes and why knowledge management programs fail to succeed, regardless of technology. In total, 95 members participated: 39 were from Gartner Customer Service & Support Research Circle and 56 were from an external survey link shared via social channels and analyst contacts. Members from North America (n = 61), EMEA (n = 20), Asia/Pacific (n = 7) and Latin America (n = 2) responded to the survey. Gartner's Customer Service and Support Research Circle members represent a mix of industries and organization sizes, with the majority in North America.

Disclaimer: Results of this survey do not represent global findings or the market as a whole, but reflect the sentiments of the respondents and companies surveyed.

[2] Large Language Models for Information Retrieval: A Survey, arXiv.

[3] In-Context Learning, arXiv.

Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

Solution Path for Knowledge Management

Developing a Knowledge Management Strategy

Quick Answer: Help Employees Find Each Other to Share Knowledge

Quick Answer: What Impact Will Generative AI Have on Search?

Prompt Engineering With Enterprise Information for LLMs and GenAI

Table 1: How Generative AI Impacts Knowledge Management

Impacts	*Top Recommendations*
Generative AI's ability to turn byproducts of normal interactions into structured knowledge assets radically reduces the effort required to capture institutional and subject matter knowledge in external, persistent forms.	■ Use large language models and Generative AI services to summarize interaction transcripts, extract key points and format the result in accordance with enterprise style guidelines and templates.
Harnessing the large language models significantly improves the productivity of knowledge workers who usually spend a significant amount of their time per day searching for information and making sense of it.	■ Transform your enterprise search by deploying LLMs to enable improved search experience and improve search accuracy.
Generative AI's capability to automatically extract data from documents and messages to produce new content promises resolution of a long-standing obstacle in knowledge management: building and maintaining expertise directories that help employees find and connect with one another for knowledge sharing and collaboration.	■ Shift to an expertise directory that uses generative AI to dynamically profile expertise by first reviewing your enterprise application portfolio, as well as what is available to buy, before developing your own solution atop an appropriate platform.

Source: Gartner (November 2023)