# GenAI unleashed

**Part 1**

# How GenAI is transforming *document search and knowledge management*

Capgemini

# From data deluge to insights

How GenAI is transforming document search and knowledge management in financial services

Organizations, particularly in the financial services sector, have long mastered the management of structured data within relational databases. These firms have honed their expertise in data storage, ensuring data quality, and leveraging this data for applications, reporting, and analytics. However, the advent of GenAI has transformed the handling of unstructured data, unlocking new possibilities in knowledge management and search capabilities across enterprise processes and workflows.

While structured data benefits from centralized storage and easy retrieval through tables and keys, managing unstructured data presents unique challenges. Ensuring that documents are not duplicated across various storage platforms like SharePoint, Teams, and Content Management Systems is less straightforward. Although some progress has been made in solving storage issues, the rigor seen in relational databases is often lacking.
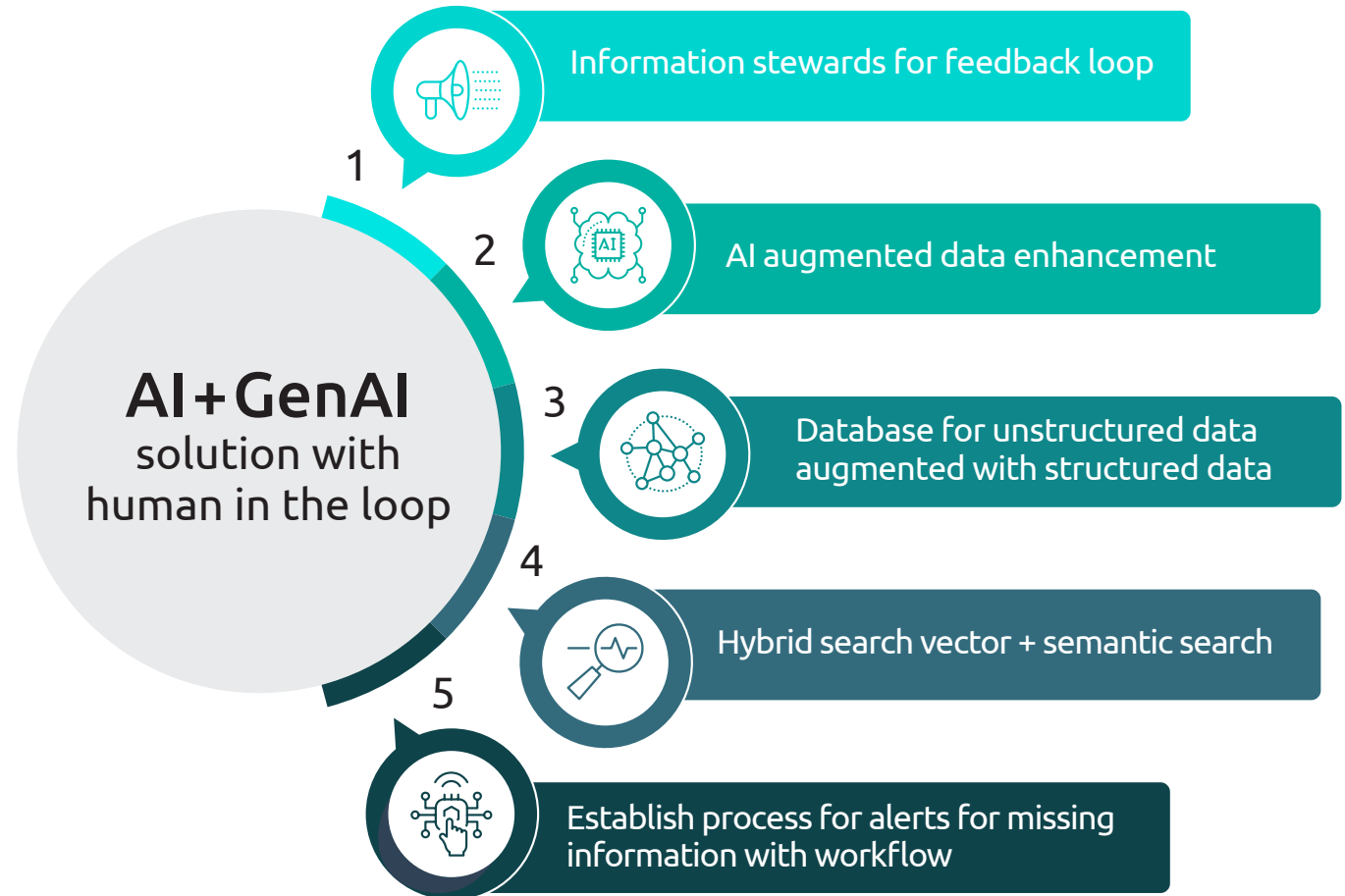
The time-consuming process of gathering and auditing information from large collections of documents can significantly hamper productivity. The complexity increases when integrating structured and unstructured data to provide a seamless and efficient user experience for business, technical, and operational purposes. The value of advanced, GenAI-powered search and knowledge management systems becomes evident, offering speed, accuracy, and scale, thus enhancing overall organizational efficiency.

# Approaching the problem from multiple fronts

Now that we've examined the challenges and business value of this organizational capability, let's discuss how to address it from multiple angles. The following chart offers an overview of the key dimensions involved in building this capability. In the subsequent sections, we will delve into the specifics of how AI and advanced techniques can be effectively implemented across the organization.

**AI+GenAI**
solution with
human in the loop

1 Information stewards for feedback loop

2 AI augmented data enhancement

3 Database for unstructured data augmented with structured data

4 Hybrid search vector + semantic search

5 Establish process for alerts for missing information with workflow
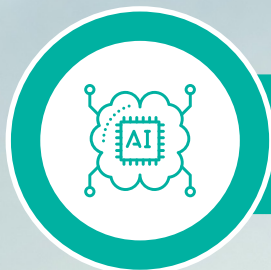
# 1) Information stewards for feedback loop

The role of Information Stewards in ensuring ongoing data readiness is crucial. Information Stewards are responsible for monitoring and managing the quality, security, and compliance of the data environment. Their oversight ensures that the data remains accurate and secure. Additionally, integrating feedback from Information Stewards is essential for continuously improving data quality and AI model performance. This ongoing process helps maintain a high standard of data readiness and enhances the effectiveness of AI implementations.

The organizational structure of the financial services firm will determine the specific responsibilities of each Information Steward. For example, every line of business (LOB) and operational horizontal, such as contact centers, back-office operations, and strategy teams, will have designated stewards. If the firm uses disparate content management systems, additional effort will be required to standardize unstructured data governance processes, ensuring the integrity of the unstructured data landscape.

## 2) AI augmented data enhancement

To ensure the quality, accuracy, and completeness of data, several capabilities are essential. Deploying classification algorithms to automatically identify document types and topics is crucial for effective data classification. Tag generation and metadata management play a significant role by automatically generating metadata tags for roles, topics, and divisions. Additionally, adhering to data standards is necessary to ensure that documents are reviewed and approved before publishing.

Document standards, such as mandatory sections for intended audience, role-based security permissions, and change audit trails, must be strictly enforced. Approaches need to be developed to automate data augmentation from system logs, incorporating this information into service desk tickets to record which systems were accessed for resolving issues. The goal is to enhance human entries with automated data from logs and other sources, thereby reducing user friction and improving the accuracy and completeness of information.

# 3) Database for unstructured data augmented with structured data

Combining structured and unstructured data involves several key strategies. Implementing vector databases for dynamic indexing of unstructured data significantly improves the speed and accuracy of search queries. Enhancing unstructured data with structured data, such as document metadata and access permissions, adds valuable context.

Adding user role-based context makes large language models (LLMs) more effective in addressing queries. By including roles and their key performance indicators (KPIs) as additional context, GenAI applications can better understand the motivations behind specific questions. This enables them to respond to general queries, such as "What are the top three things I should worry about today?" with greater expertise and relevance.

Additionally, exploring advanced techniques like combining Retrieval-Augmented Generation (RAG) architecture with knowledge graphs can further augment the enterprise context, providing a more comprehensive and efficient data management solution. GraphRAG approaches add an extra advisory layer that helps identify related document chunks specific to the document repository being queried.

## 4) Hybrid search vector + semantic search

To enable quick and effective data search and presentation to end users, a hybrid search and agentic architecture is essential. This approach combines the precision of vector search with semantic search to enhance search accuracy. Result enhancement is achieved through ranking fusion techniques, which merge results from both search types.

Additionally, the ability to call APIs across multiple domains, such as CRM, document repositories, service desks, and requirements, further enriches the search capabilities. An agentic architecture, with libraries for specific functionalities, ensures an improved customer experience (CX). This architecture allows AI libraries to augment GenAI applications' capabilities, such as performing mathematical calculations, rendering reports, and creating SQL queries against specific databases.

This evolution is crucial as it enables applications to explore areas like intelligent decision-making, rules execution, and product recommendations. The goal is twofold: first, to enhance enterprise context retrieval, and second, to augment GenAI with AI and other APIs to deliver a superior customer experience.

# 5) Establish process for alerts for missing information with workflow

To automate continuous monitoring of processes and workflows, it's essential to integrate systems for alerts and monitoring. Establishing a monitoring and alerts system allows for the oversight of data quality and completeness, promptly notifying teams of any anomalies or gaps.

Once alerts are triggered, workflow automation is used to respond efficiently, with predefined workflows in place to address and rectify identified data issues. This ensures timely and effective resolution of data quality problems.

Given that this is an ongoing effort, there is a pressing organizational need to keep the data fresh, up-to-date, and complete with the highest level of quality. This dedication to data integrity ensures that users receive the best possible information when they need it.

# Bringing it all together

While financial services firms have long excelled in managing structured data within relational databases, the advent of GenAI has opened up transformative possibilities for handling unstructured data. This evolution is crucial for enhancing knowledge management and search capabilities across enterprise processes and workflows. Managing unstructured data poses unique challenges, including preventing document duplication across various storage platforms and ensuring data accuracy and completeness.

**To overcome these challenges, the problem needs to be approached from multiple fronts:**

- Information Stewards ensure data quality, security, compliance, and continuous improvement of AI performance.
- Classification algorithms and metadata management ensure data quality and adherence to standards.
- Combining structured and unstructured data with vector databases and RAG architecture improves search accuracy.
- Incorporating hybrid vector and semantic search, ranking fusion, and API integration further refines search precision.
- Monitoring and alert systems with automated workflows maintain data quality and completeness.

By addressing these challenges from multiple fronts and leveraging advanced AI techniques, financial services firms can unlock the full potential of their data, driving superior decision-making and operational efficiency.

# Meet our experts

## Shankar Ramanathan
Senior Director
AI & Machine Learning
Financial Services Insights & Data
shankar.a.ramanathan@capgemini.com

in

## Rajesh Iyer
Global Head of AI and ML
Financial Services
rajesh.iyer@capgemini.com

in

## Vishal Bhalla
Senior Director
Portfolio Lead
Financial Services Insights & Data
vishal.bhalla@capgemini.com

in

**Data-powered financial services**
Learn how we can help your organization harness
the power of data using AI, machine learning,
process automation, and more.

## About Capgemini

Capgemini is a global business and technology transformation partner, helping organizations to accelerate their dual transition to a digital and sustainable world, while creating tangible impact for enterprises and society. It is a responsible and diverse group of 340,000 team members in more than 50 countries. With its strong over 55-year heritage, Capgemini is trusted by its clients to unlock the value of technology to address the entire breadth of their business needs. It delivers end-to-end services and solutions leveraging strengths from strategy and design to engineering, all fueled by its market leading capabilities in AI, cloud and data, combined with its deep industry expertise and partner ecosystem. The Group reported 2023 global revenues of €22.5 billion.

**Get the Future You Want | www.capgemini.com**

Capgemini