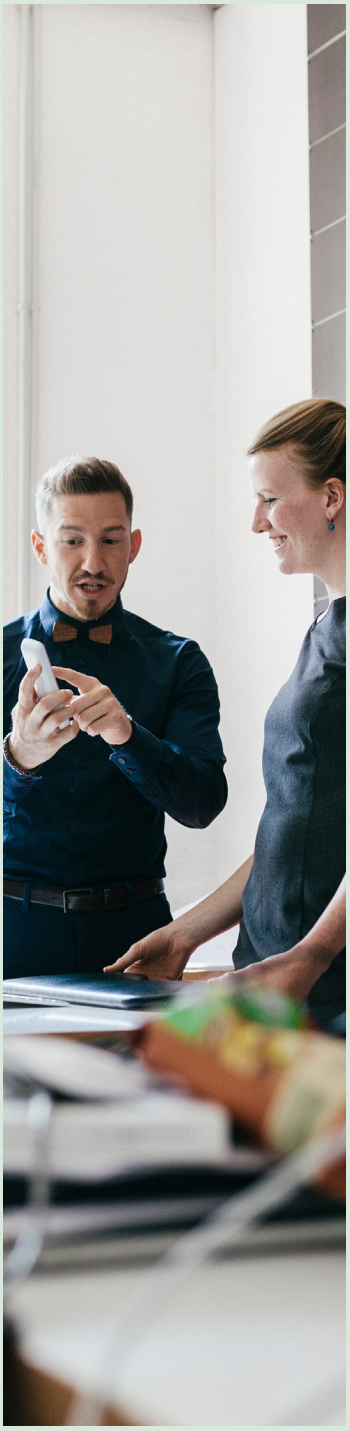
WHITE PAPER



GENAI DIGITAL KNOWLEDGE COWORKERS POWERED BY RAG TECHNIQUE WITH TRUSTED SOURCE

TRANSFORMING DIGITAL KNOWLEDGE MANAGEMENT

ABSTRACT

The future of Digital Knowledge Management lies in the integration of secure and insightful Digital Knowledge Coworkers powered by Large Language Models (LLMs) with the Retrieval-Augmented Generation (RAG) technique which sources information from a well-controlled, and governed document management system. This White Paper showcases how this combination promises enhanced knowledge retrieval, insightful analysis and continuous learning. Thus, revolutionizing the way organizations manage and use their knowledge assets.

Infosys®
Navigate your next

Introduction

The dynamics of what's #NextAtWork are changing at an unprecedented pace. Especially, with the advent of GenAI (Artificial Intelligence), the critical role of Knowledge Management (KM) has been elevated to make a positive impact on all three dimensions – work, workspace and workforce. More so, KM will become the cornerstone that supports organizational resilience, enables businesses to be adaptable to change faster and gain a competitive edge to stay ahead.

Now, imagine coworkers who never take leaves, never retire, whose knowledge is ever expanding, and who always have the latest information at their fingertips. Coworkers who are available to work – anywhere, anytime. Furthermore, coworkers on whom everyone in the organization can rely on to obtain a right summary and request actions without having to leave the flow of work or switch context. This is the future of Knowledge Management with the backing of Digital Knowledge Coworkers in this era of AI. Therefore, Digital Knowledge Coworkers will ensure correct and relevant information is available to the right person at the right time.

Eventually, there will be AI infused products and platforms which will improve everyday individual productivity. However, what will take the flow of work to a new level are the transformational AI products. One of them is a transformational AI infused solution - a Digital Knowledge Coworker that offers to bring quality information to the fingertips, provide actionable insights and completes the human approved actions. This paper focuses on how to deploy such a Digital Knowledge Coworker iteratively in a practical way to reap the benefits of GenAI and RAG (Retrieval Augmented Generation).

The figure illustrates how a GenAI Digital Knowledge Coworker relies on up-to-date information from a trusted source or document management system.

Note: the quality of the output is only as good as the quality of the data source provided and the controls laid in the foundation model.

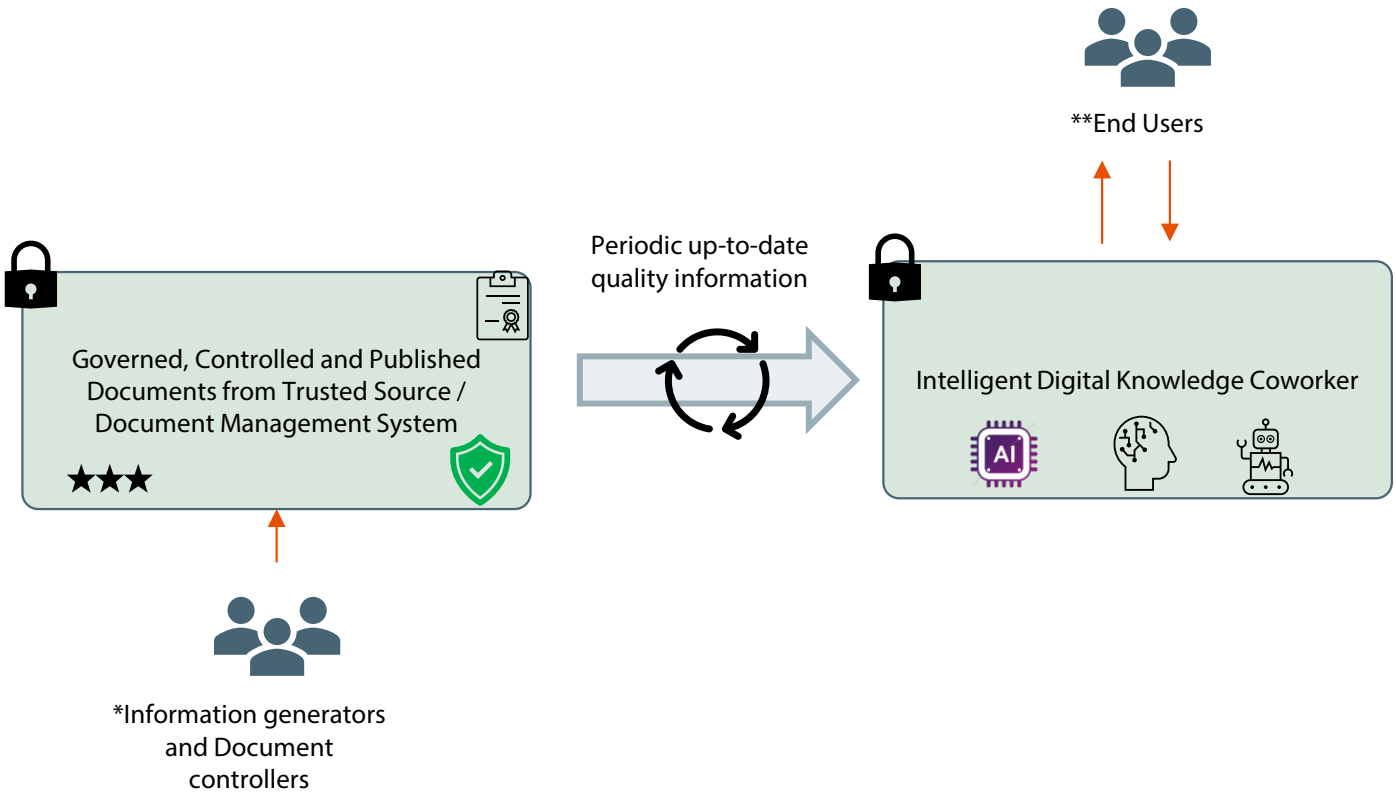


Figure 1 : GenAI Digital Knowledge Coworker Needs Quality Data Inputs

Digital Knowledge Coworker and its Role in Knowledge Management

In the present AI era, this Digital Knowledge Coworker is fast becoming a reality. Especially, for businesses looking for smarter ways to leverage AI to unlock productivity, amplify human-centric experiences and democratize innovation to shape the future of work.

With the advent of these GenAI platforms for the enterprises, the quality of output is only as good as the information these new era platforms are fed with. Not to forget the risks related to hallucinations and bias that could be induced in this process.

At this point, GenAI Platforms neither replace search nor machine learning algorithms, rather it augments them and exponentially enhances productivity by deducing the key insights. While the advantages are quite rewarding, the implementation of this next generation Coworker platform comes with its own set of challenges. Currently, most large enterprises are facing deep quality issues in the results surfaced from simple search engine or an enterprise search engine due to lack of governance on the underlying knowledge source.

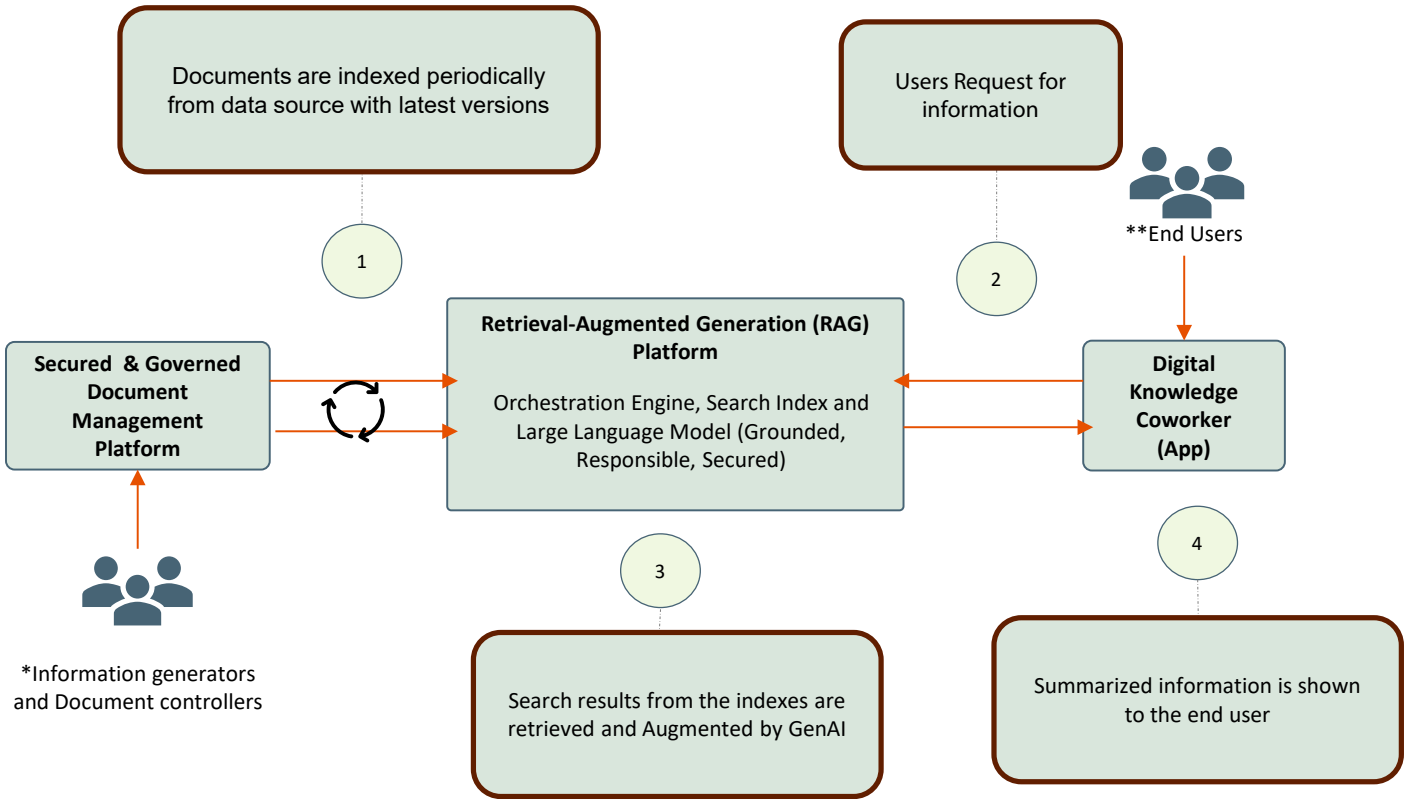


Figure 2: Conceptual Overview of the Digital Knowledge Coworker

In this pursuit as shown in the above figure, a successful Digital Knowledge Coworker requires:

1. Generative AI (GenAI) powered by RAG Technique
2. Secured & Governed Document Management Platform

When GenAI is backed by groundbreaking RAG (Retrieval Augmented Generation), this technique can make a profound impact. RAG platform gets the top search results from data sources and provides the summary based on the instructions or prompts provided by the end user. Also, when RAG Platform is anchored to governed, organized and vetted enterprise data with safe, responsible, and secure modules, this platform will become a true Digital Knowledge Coworker. While the information can be from various sources (product documentation, support tickets, historical data, and training materials, among others), the most important part one needs to be concerned about is that the underlying information source is well governed and controlled. That said, this document management platform provides up to date information, which is approved for publishing, or viewing. Further, it disposes or archives expired content and uses the right permissions to display the information.

Use Cases and Themes

Here are some common use cases and themes across industries positively impacted by the coming together of GenAI and RAG:

A Comparison - Traditional Experience vs. New GenAI Powered Experience

Here are some common use cases and themes across industries positively impacted by the coming together of GenAI and RAG:

Key Themes	Traditional Experience	New Experience with GenAI Powered by RAG
Search	Primarily manual way of finding information from one document link to another	A summarized information from various documents made available at once, with references and thought process of how the knowledge was obtained
Completeness of information	If the information is spread across various documents, there is a high probability for the user to miss parts of the information	With the vectorization of the information and hybrid search option, the user will have access to maximum information within the limits of the platforms
Results quality & accuracy	The top search results depend on the quality of the search engine and quality of the information available	Practices like grounding, fine tuning, Responsible AI and vectorization have direct impact on improving the quality of information
Platform scalability	Search platform provides the base RAG Technique	Grounding** this information from these data sources / search results only can make this platform trustworthy and scalable

New Experience with GenAI Powered by RAG

Contrary to the Traditional User Experience, here the Digital Knowledge Coworker is only connected to specific controlled and governed data sources, so the output is grounded as shown in the flow below:

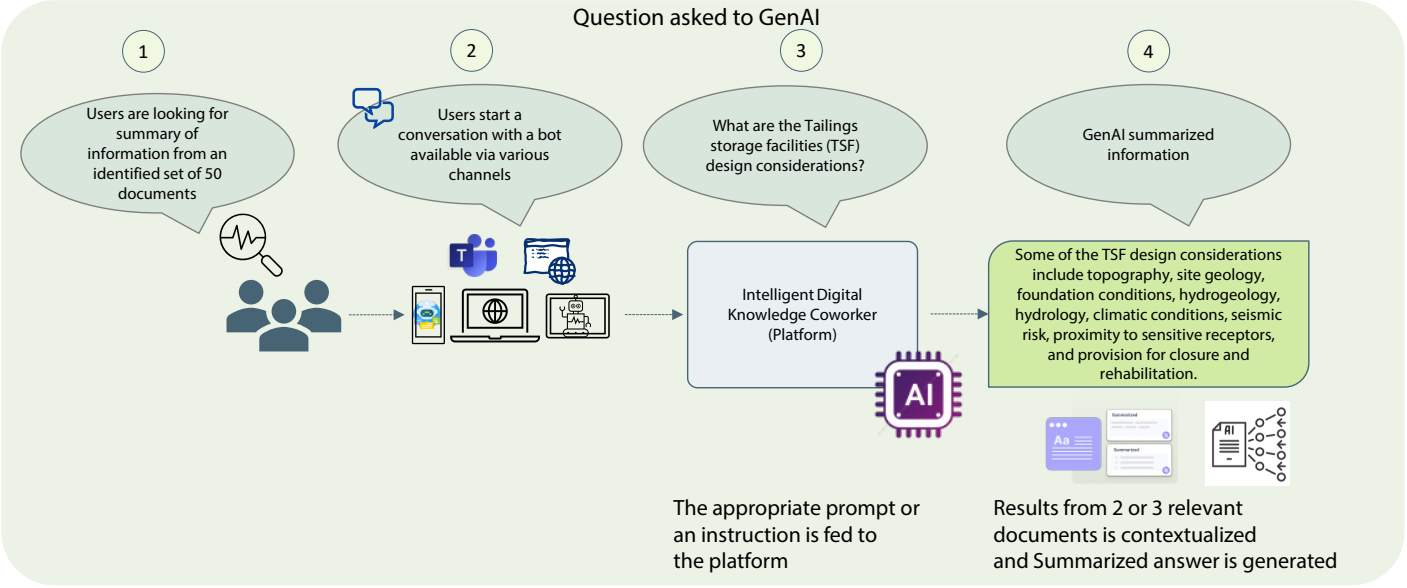


Figure 5: New user experience with GenAI

The user requests for safety information → The Digital Knowledge Coworker gets all results from search service and passes the results along the right prompts to the Large Language Model → The result is a summarization of the exact safety procedures the user was looking for rather than the links which he would go investigate one by one to get the relevant information

About GenAI Grounding**

When the prompts or models of the LLMs (Large Language Models) are anchored or restricted to use only the specific data sources or use cases, they produce exact and relevant summarization. This anchors the model to specific information. The following flow articulates the scenario:

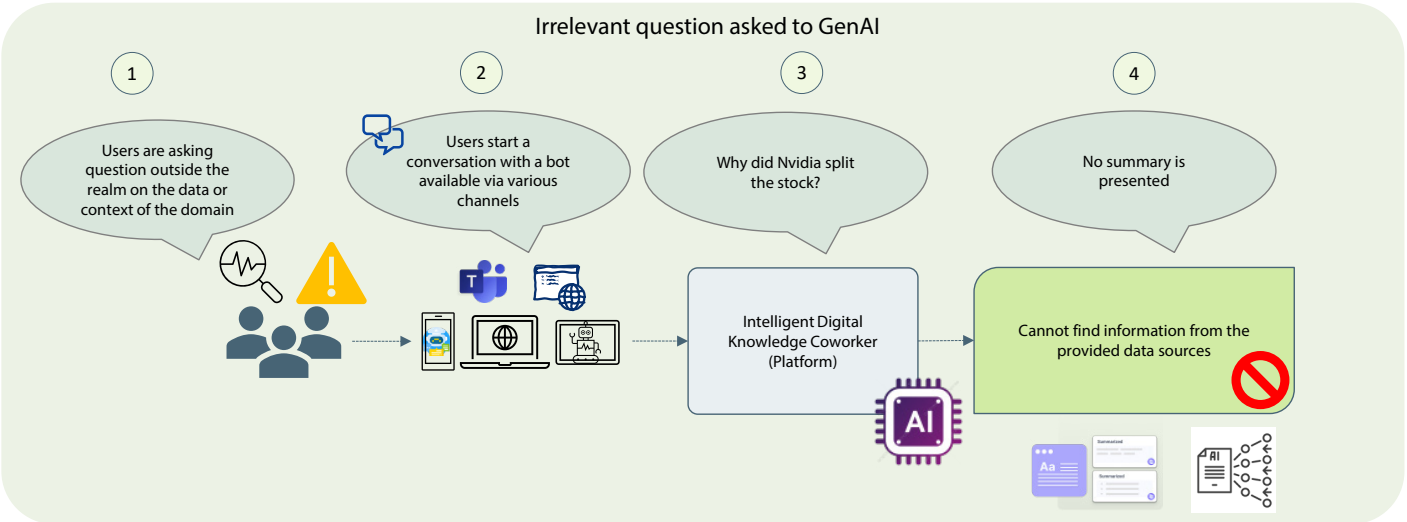
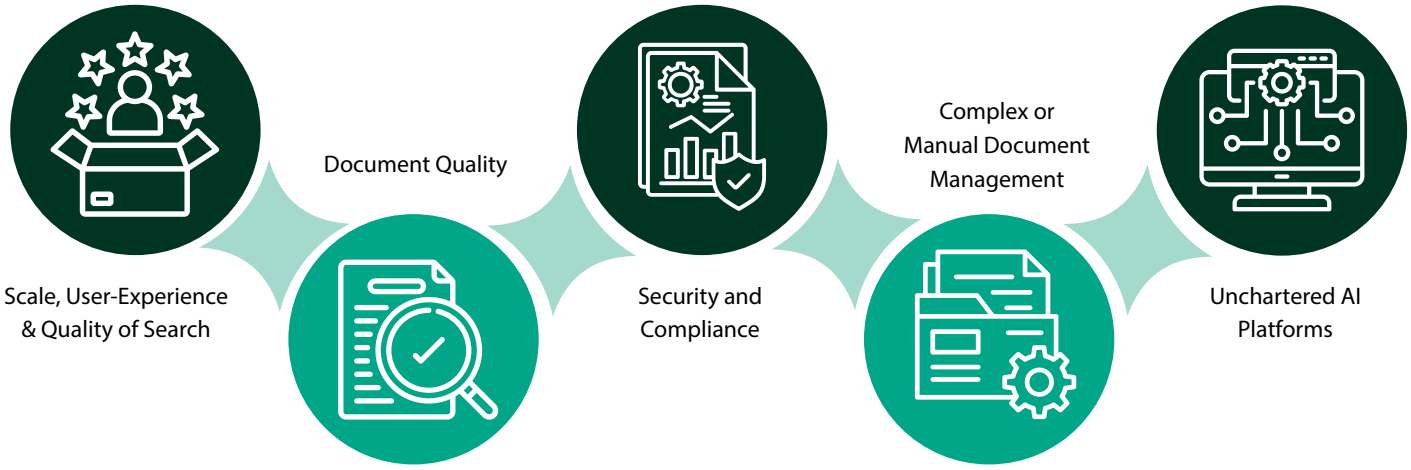


Figure 6: GenAI Grounding

The user requests information which is outside the realm of the data or context of the Digital Knowledge Coworker → The LLM (Large Language Model) is not associated or connected with any Data Source → The result is – no information found. This is due to the anchoring or grounding of the Large Language Model

Challenges in Deploying Digital Knowledge Coworker

Contrary to the Traditional User Experience, here the Digital Knowledge Coworker is only connected to specific controlled and governed data sources, so the output is grounded as shown in the flow below:



1. Scale, User-Experience & Quality of Search

The challenges primarily stem from the exponential growth of unstructured data collected by companies. Lack of accurate metadata, absence of data management solutions, a poor understanding of search user behavior, inadequate employee training on the effective use of search tools amplifies the challenges. Lastly, user adoption, an essential aspect for the success of any tool or process, is also a challenge.

2. Document Quality

The first set of challenges includes inaccuracy, outdatedness, and duplicity. That apart, data incompleteness, and cross-system inconsistency, which occurs when data or documents don't align across different systems. Lack of standardization, and version control issues, where it's difficult to manage and find different versions of a document or data set.

3. Security and Compliance

One of the concerns is user access limitations, which refers to the restriction of user access to certain data or documents based on their role or level in the organization. Legacy systems, not equipped with the latest security measures pose a risk. Accidental data exposure, insufficient personnel or technology to adequately protect data, cyber threats, such as hacking and phishing attacks, privacy concerns are also some of the major challenges faced by organizations. Lastly, failure to adhere to legal and industry standards for compliance and regulations can result in penalties and damage to the organization's reputation.

4. Complex or Manual Document Management

One of the challenges is the reliance on paper-based documents. This traditional documentation method is environmentally unfriendly, space consuming and inefficient in storage, modification and accessibility, and prone to damage or loss. This problem is worsened when the documents are not meticulously organized or indexed. Lastly, complex workflows pose a challenge in document management. These workflows may involve multiple stages of review, approval, and modification, which can be difficult to track and manage effectively. The use of digital document management systems can help address these challenges by providing features like advanced search, version control, and workflow automation.

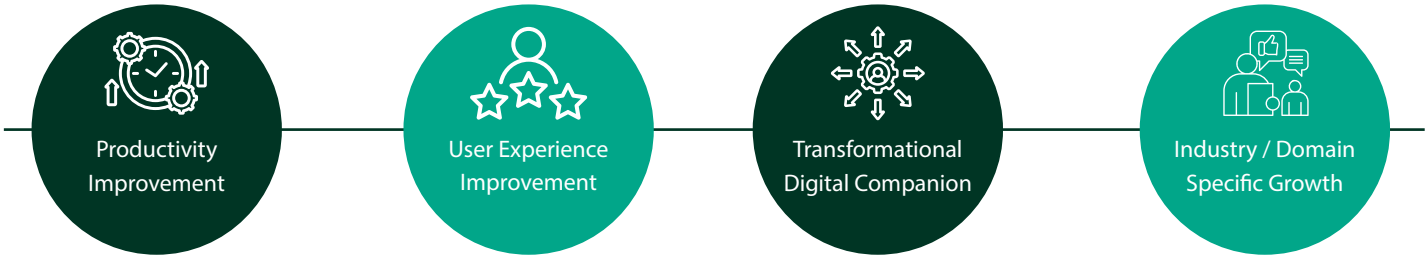
5. Unchartered AI Platforms

GenAI adoption presents challenges, particularly concerning data sensitivity and intellectual property (IP). Another significant concern is the potential for GenAI models to inherit biases from their training data, which can lead to skewed or unfair outcomes. The creation of content by these systems may also have legal implications, especially if it infringes on existing laws or regulations. Additionally, understanding copyright and patent law in the context of GenAI is crucial. Lastly, GenAI systems can sometimes produce "hallucinations" or generate content that is not based on the input data.

That said, these challenges can be solved by the right governance structure, security configuration and better user experience. Let us discuss the benefits of this Digital Knowledge Coworker when deployed and how much it would be beneficial for the employees.

The Benefits of Digital Knowledge Coworker

Contrary to the traditional User-Experience, here the Digital Knowledge Coworker is only connected to specific controlled and governed data sources, so the output is grounded as shown in the flow below:



1. Productivity Improvement

By 2030, it is projected that the productivity will increase fourfold. Application modernization is also expected to improve by 30%, and human resources functions are set to see a 40% improvement. Interestingly, this productivity improvement is predicted to influence 4.5 times more jobs than it replaces, showing its potential for significant positive impact on the job market.

2. User Experience Improvement

The main goals are to enhance the user experience, increase customer satisfaction by 70%, and improve existing business processes.

3. Transformational Digital Companion

Digital Knowledge Coworkers are not just tools but are becoming more personal and familiar, thus aiming to enhance productivity and provide continuous support, any given time and anywhere.

4. Industry/ Domain Specific Growth

Manufacturing

- By using data analysis for decision-making improving efficiency
- Understanding equipment telemetry data
- Enhancements in design processes are reducing time to market
- Employee training and information management
- Predictive maintenance is being achieved through historical data analysis
- Quality control is being enhanced by anomaly detection.
- Regulatory compliance is being managed through automated document processing and knowledge management is helping in handling vast amounts of data

Mining

- Using predictive analytics to manage supply chain disruptions, keep perfect inventory levels and find potential disruptions
- Environmental data analysis is being used to reduce environmental risks and impacts
- Exploration process is being improved by analyzing massive amounts of data

Energy

- Equipment management is a key growth aspect, ensuring the best operation and longevity of equipment
- Health and safety measures prioritized to ensure a safe working environment

Healthcare

- Seeing growth in areas like automated or assisted diagnosis and prescription, auditing prescriptions, real-time prioritization, and triage
- Personalized medication and care are being provided based on patient-specific data.
- Patient data analytics, benefits verification, and auto rules generation are also being used to improve healthcare services

Conclusion – The Way Forward

Organizations already invested in enterprise search platforms or domain specific tools can enable Digital Knowledge Coworkers with GenAI with RAG without having to rebuild a new infrastructure for GenAI transformation. If these search platforms are not as mature, then they should consider to progressively build with the RAG linked GenAI Digital Knowledge Coworkers enabled by a strong search foundation. Moreover, this foundational block will ensure accuracy, show recency of the documents with efficient document management and facilitate easy information retrieval across the organization. Once this quality of information is baselined, the transformational GenAI use cases of a true autonomous bots will become a reality.

About the Author

Siraj Shaik

Principal Technology Architect, Digital Workplace Services, Infosys

Siraj is a Digital Workplace Ecosystem Transformation Leader & Enterprise Architect. Also, he is a Certified Azure Solutions Architect & TOGAF Certified Enterprise Architect and is helping clients implement GenAI RAG solutions.

For the last 20+ years, Siraj has been working on Enterprise Architecture and Solution Architecture roles of Modern Apps, Azure/AWS IaaS, PaaS, Digital Workplace, Portals and Content Management Solutions using Microsoft Suite of products.

For more information, contact askus@infosys.com

Infosys®
Navigate your next