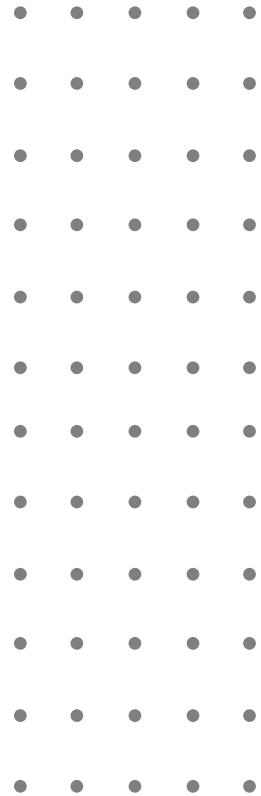




# Enterprise data readiness for GenAI in 2024

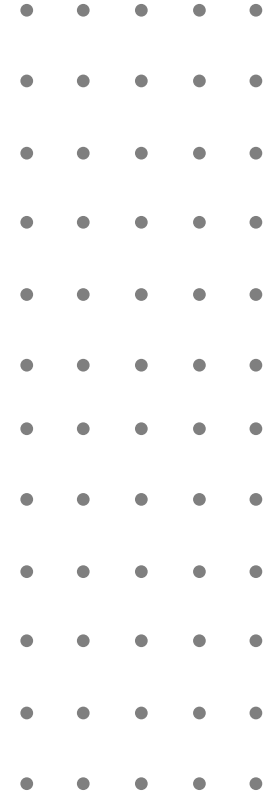
October 2024



## Table of Contents

|                                                                                      |           |
|--------------------------------------------------------------------------------------|-----------|
| <b>Introduction and key findings.....</b>                                            | <b>3</b>  |
| <b>The state of GenAI adoption .....</b>                                             | <b>7</b>  |
| Most organizations are moving beyond exploration to pilots with GenAI.....           | 8         |
| Top GenAI use cases for the next 12 months.....                                      | 9         |
| Only 14% of organizations are using generic LLMs .....                               | 10        |
| Status of RAG adoption .....                                                         | 11        |
| Most popular LLMs in use today .....                                                 | 12        |
| <b>Top concerns .....</b>                                                            | <b>13</b> |
| Top challenges in deploying GenAI applications .....                                 | 14        |
| Top concerns in leveraging enterprise data for GenAI and RAG (all respondents) ..... | 15        |
| Top concerns in leveraging enterprise data for GenAI and RAG (by industry) .....     | 16        |
| <b>Data-readiness .....</b>                                                          | <b>17</b> |
| Most valuable internal data sources for customer-facing GenAI .....                  | 18        |
| Top systems housing customer data .....                                              | 19        |
| Types of internal data used for GenAI and RAG projects .....                         | 20        |
| <b>Demographics .....</b>                                                            | <b>21</b> |
| <b>About K2view .....</b>                                                            | <b>24</b> |

# Introduction and key findings



## Introduction & methodology

---

The enterprise data landscape is undergoing rapid transformation, largely driven by the advent of Generative AI (GenAI). As organizations increasingly explore GenAI's potential, many are encountering significant challenges related to their existing data infrastructure. These challenges, once manageable or hidden, are now being exposed and exacerbated by the demands of GenAI, particularly in areas like data security and data readiness.

As [Gartner](#) predicts, by the end of 2025, 30% of GenAI projects will be abandoned after proof of concept due to poor data quality, inadequate risk controls, escalating costs, and unclear business value. This underscores the critical role of data in the success of GenAI initiatives.

We conducted this survey to understand the specific obstacles organizations face in preparing their enterprise data for GenAI projects. Our focus is on how these challenges manifest, how they impact the transition from pilot to production, and what strategies organizations are employing to overcome them. Our survey delves into these issues, examining how organizations are navigating the GenAI data gap—a gap characterized by the inability to effectively leverage vast amounts of data spread across various systems for customer-centric GenAI projects.

Through this survey, we aim to shed light on the current state of data readiness for GenAI within enterprises, identify the obstacles hindering progress, and highlight the solutions that are being deployed to ensure that GenAI projects not only reach production but deliver sustained value. Our findings will be particularly relevant to those leading or participating in GenAI projects, providing insights into best practices and strategies for overcoming data-related challenges in this rapidly evolving field.

### Methodology

This survey was conducted in July 2024 by Global Surveyz, an independent survey company. We gathered insights from 300 senior professionals who are directly involved in the planning, building, or delivery of GenAI applications. The respondents were evenly split between IT and Data departments, all holding Director-level positions or higher.

Participants were drawn from B2C industries, including Banking, Insurance, Healthcare, Retail, and Telecommunications, with a balanced representation across these sectors. The survey covered companies with over 1,000 employees, primarily based in the United States (70%), with the remainder in the United Kingdom (30%).

Our survey focused exclusively on professionals actively engaged with GenAI applications, ensuring that the insights reflect the real-world challenges and strategies of those on the front lines of this evolving technology.

## Key findings

---

### 1 **The biggest obstacles to deploying GenAI applications all relate to enterprise data.**

While cost is always a major concern (51%), the next top challenges in implementing GenAI applications are all related to data: security and privacy (48%), the reliability of large language models (44%), and the readiness of enterprise data for GenAI (33%). These issues highlight the critical need for secure and reliable data management in the deployment of GenAI technologies.

### 2 **Retrieval Augmented Generation (RAG) is already widely adopted as a leading framework for grounding Large Language Models (LLMs) with internal data, used by 72% of organizations.**

Only 14% of organizations are using generic LLMs. 60% are currently piloting RAG as a leading framework to ground LLMs with internal data, and 12% have already moved it into production, indicating RAG's strong position in the GenAI landscape.

### 3 **Organizations require both unstructured and structured data to ground their GenAI applications.**

The survey reveals that respondents are using both unstructured data sources such as vector databases (44%) and graph databases (38%), as well as structured data sources like operational systems (54%) and data lakes (45%). This variety of data sources underscores the complexity of integrating enterprise data for GenAI and RAG projects.

4

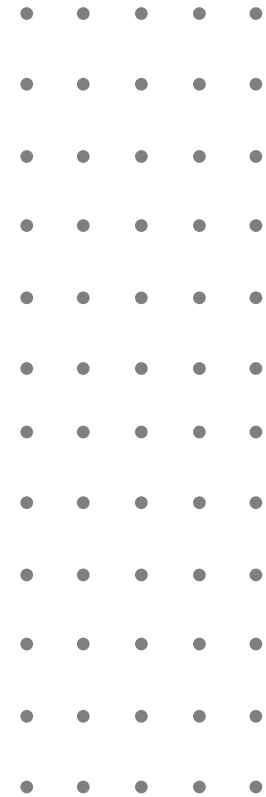
#### **The top challenges in using enterprise data for RAG are scalability, data quality, real-time access, and data protection.**

The survey reveals a unique set of challenges specific to enterprise application data, including scalability and performance (48%), data quality and consistency (46%), real-time access (46%), data governance and compliance (44%), and data security and privacy (43%). Organizations have yet to fully address these unique challenges when it comes to managing and leveraging enterprise data for GenAI and RAG initiatives.

5

#### **Most of the data required to augment customer-centric GenAI applications comes from structured enterprise data that resides in many applications.**

The survey indicates that most of the value derived from customer-centric GenAI applications comes from structured enterprise data fragmented across various systems, including CRM (57%), finance (53%), and ERP (40%). The large fragmentation poses a significant data integration and data governance issue in leveraging the data for GenAI.



# The state of GenAI adoption

## Most organizations are moving beyond exploration to pilots with GenAI

As GenAI rapidly evolves, organizations are at various stages of experimenting, piloting, and deploying this transformative technology. We asked survey respondents how their companies are currently addressing GenAI and the extent to which they have integrated it into their operations.

43% of respondents report being involved in department or business unit-level pilots, and 13% are conducting organization-wide pilots. Only a small fraction—2%—have advanced to the production deployment of one or more GenAI use cases. An additional 17% are engaged in experimentation at the business unit level and 21% of companies are conducting organization-wide experimentation.

The results show that most organizations are moving beyond the exploratory phases into pilot projects. This aligns with broader industry trends, where companies are actively experimenting and conducting pilot projects to assess the feasibility and value of GenAI before committing to broader deployment.

This finding highlights the cautious, yet proactive approach businesses are taking as they navigate the complexities and opportunities presented by GenAI.

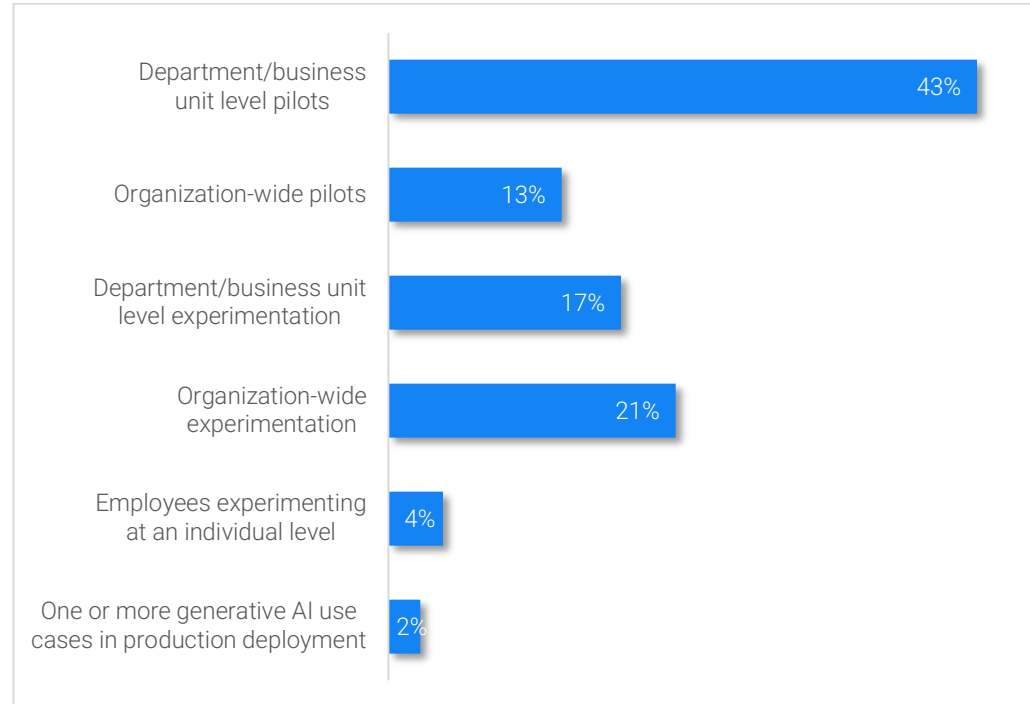


Figure 1: Current approaches to GenAI adoption

## Top GenAI use cases for the next 12 months

Our survey identified the top areas where companies plan to deploy GenAI solutions.

Leading the charge, 63% of respondents indicated that their organizations are planning to implement GenAI in marketing and sales within the next 12 months. Closely following are customer operations, with 54% of companies prioritizing GenAI use cases in this area that includes customer support through AI-driven chatbots and agent assistance.

A deeper look into the data reveals that 83% of respondents from the Retail and Telco sectors are planning to implement GenAI in customer operations, followed by Travel & Hospitality (74%) and Financial Services (47%).

These findings indicate that while marketing and sales remain central to GenAI adoption, customer operations are rapidly emerging as a critical focus area. Organizations across industries are recognizing the value of GenAI in improving customer engagement and satisfaction, reducing response times, improving first contact resolution, and increasing cross and upselling.

Question allowed more than one answer and as a result, percentages will add up to more than 100%

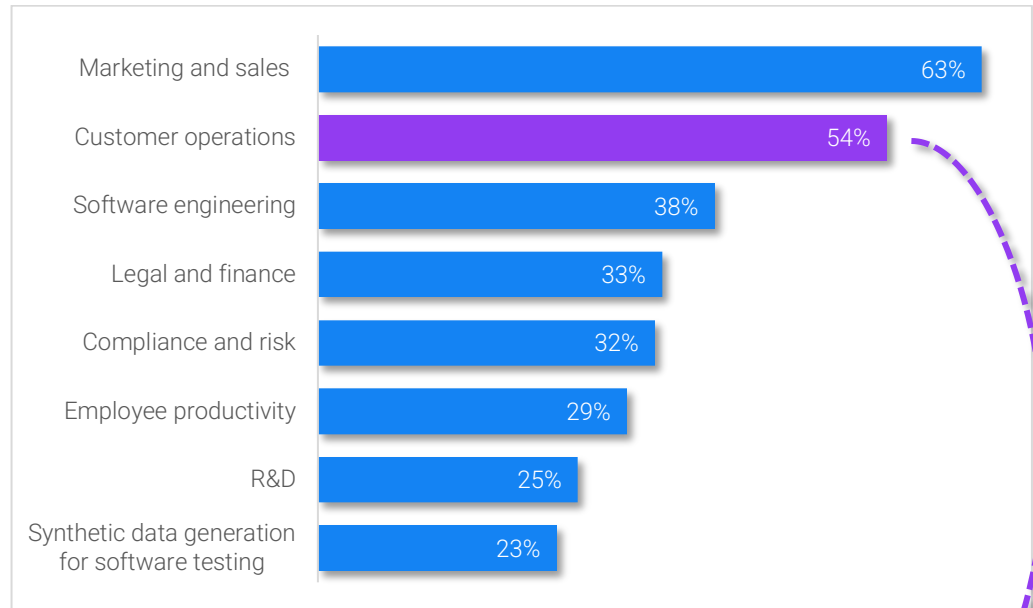


Figure 2: Top GenAI use cases for the next 12 months

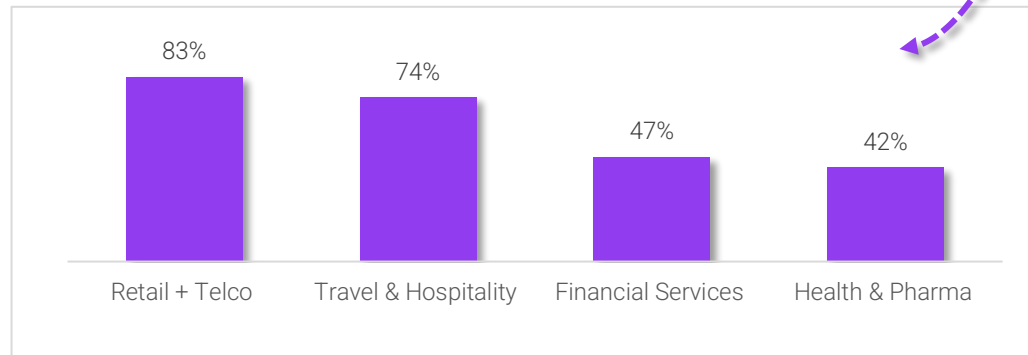


Figure 3: GenAI for customer operations use case, by industry

## Only 14% of organizations are using generic LLMs

Only 14% of organizations are using generic Large Language Models (LLMs) in their AI projects. The overwhelming majority—86%—are opting to augment their LLMs, recognizing that out-of-the-box models often lack the customization needed to meet specific business needs.

Techniques like RAG, fine-tuning, and embedding to enhance their LLMs allow organizations to infuse their AI models with relevant internal data, improving performance and accuracy for customer-facing and operational applications.

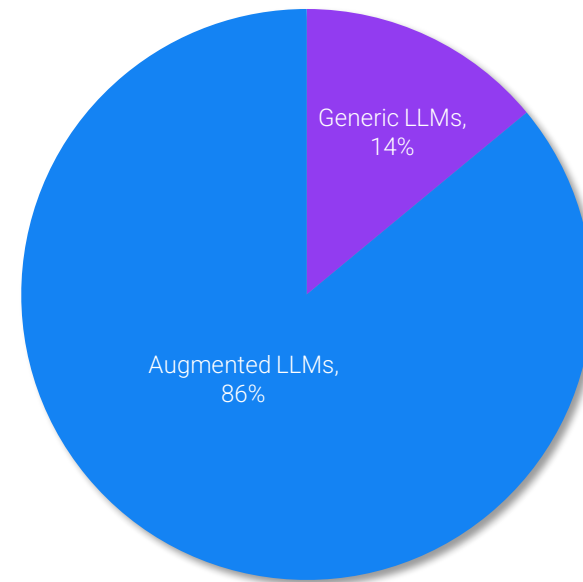


Figure 4: Leading LLM approaches in use today

## Status of RAG adoption

The adoption of RAG is rapidly progressing, with most organizations actively piloting this technology. Our survey shows that 60% of respondents are already piloting RAG, while 24% are in the planning stages, and 12% have moved to full production. A small percentage (4%) are still in the exploratory phase.

Industry-specific data shows that Health & Pharma leads in RAG pilots at 75%, followed by Banking and Financial Services at 61%, Retail & Telco at 52%, and Travel & Hospitality at 29%.

These results indicate strong momentum in adopting RAG, particularly in industries where data privacy and response accuracy are critical. However, while piloting is widespread, the transition to full production is very limited, reflecting the complexities of scaling these technologies.

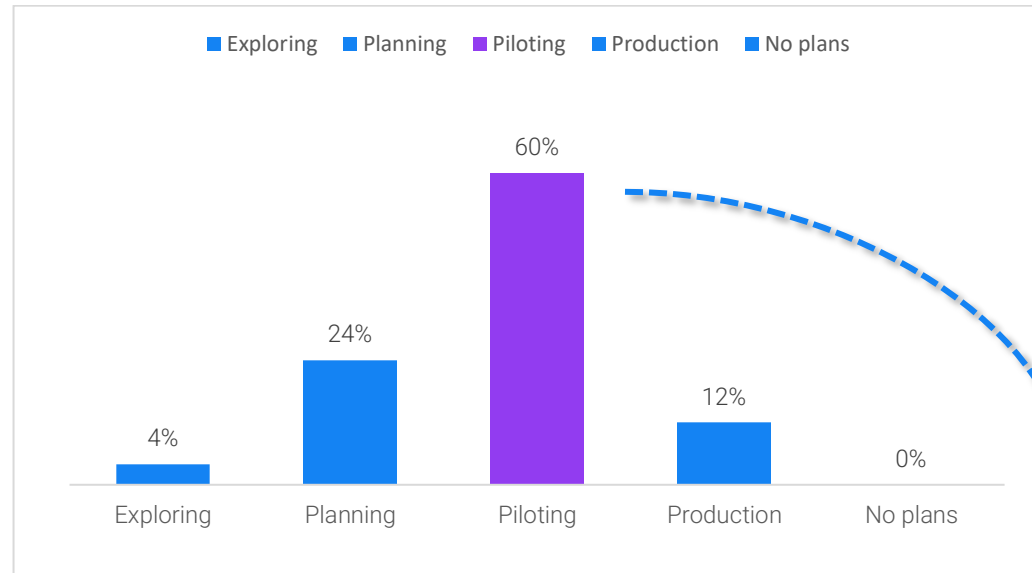


Figure 5: Status of RAG adoption

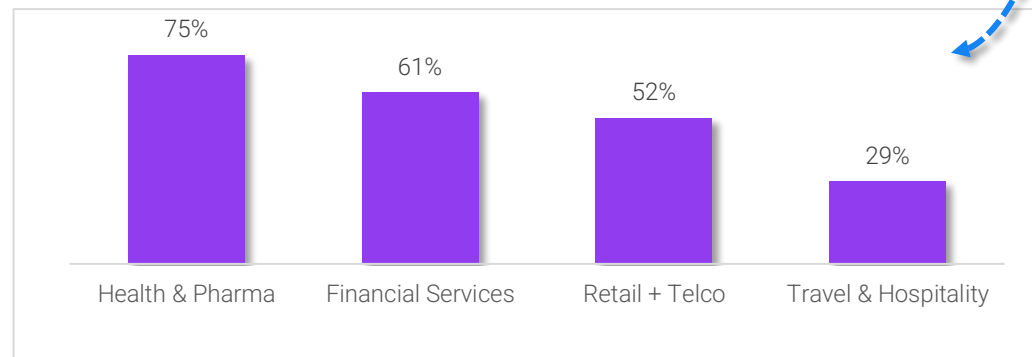
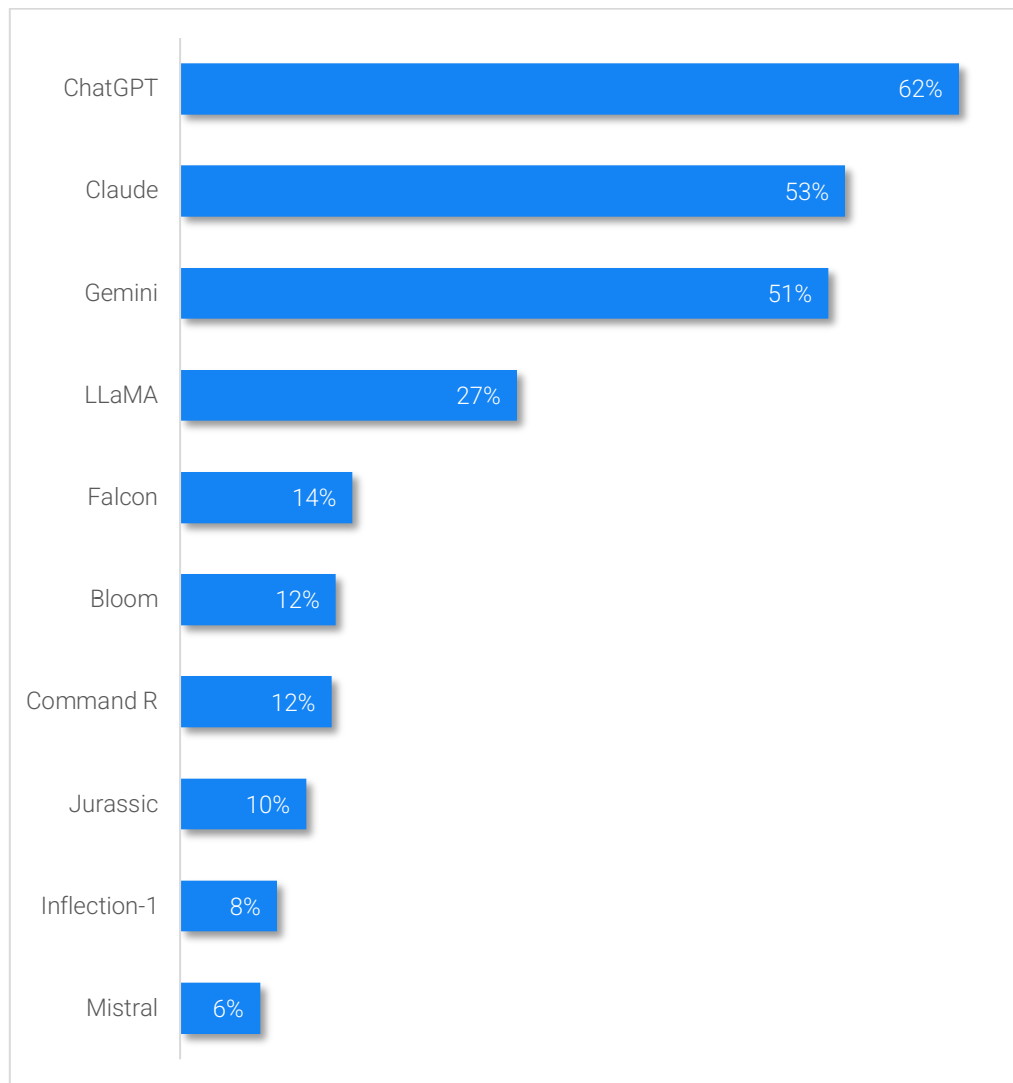


Figure 6: RAG on piloting stage, by industry

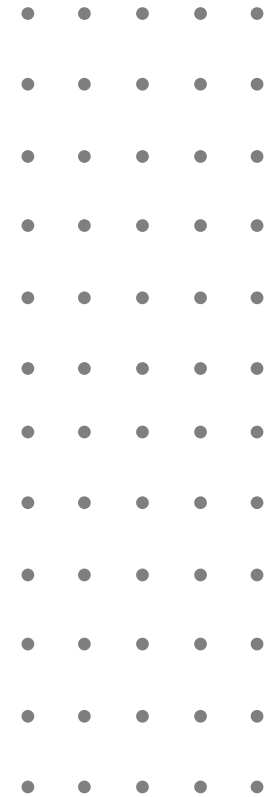
## Most popular LLMs in use today

A very broad spectrum of LLMs is used to power enterprise GenAI apps. Our survey reveals that OpenAI's ChatGPT is the most widely used LLM, with 62% of respondents reporting its adoption within their organizations. Following closely are Anthropic's Claude, used by 53% of respondents and Google's Gemini, adopted by 51%. Meta's LLaMA is also gaining traction at 27%.



Question allowed more than one answer and as a result, percentages will add up to more than 100%

Figure 7: Most popular LLMs in use today



# Top concerns

## Top challenges in deploying GenAI applications

The top obstacle in deploying GenAI applications is cost, chosen by 51% of survey respondents. This reflects the cost of the LLMs (tokens), personnel, data infrastructure, and building and testing GenAI apps—a lot of unknowns that are hard to predict and control.

Close behind are concerns about data security and privacy (48%); it's crucial to ensure that sensitive data is securely protected, preventing any risk of it being exposed or inadvertently shared through the LLM.

Reliability of LLM responses (44%) is the third major concern. Organizations need to trust that the AI-generated outputs are accurate, personalized, and reliable enough to be shared with customers.

Additionally, 33% of respondents identified enterprise data readiness as a significant hurdle. This challenge is more pronounced in larger companies, with 45% of them citing it as an issue, compared to 29% of smaller companies. The difficulty lies in the fragmented nature of enterprise data, which is often spread across multiple systems and formats, making it hard to integrate, govern, and make accessible in near-real time to GenAI applications.

\*\*Question allowed more than one answer and as a result, percentages will add up to more than 100%

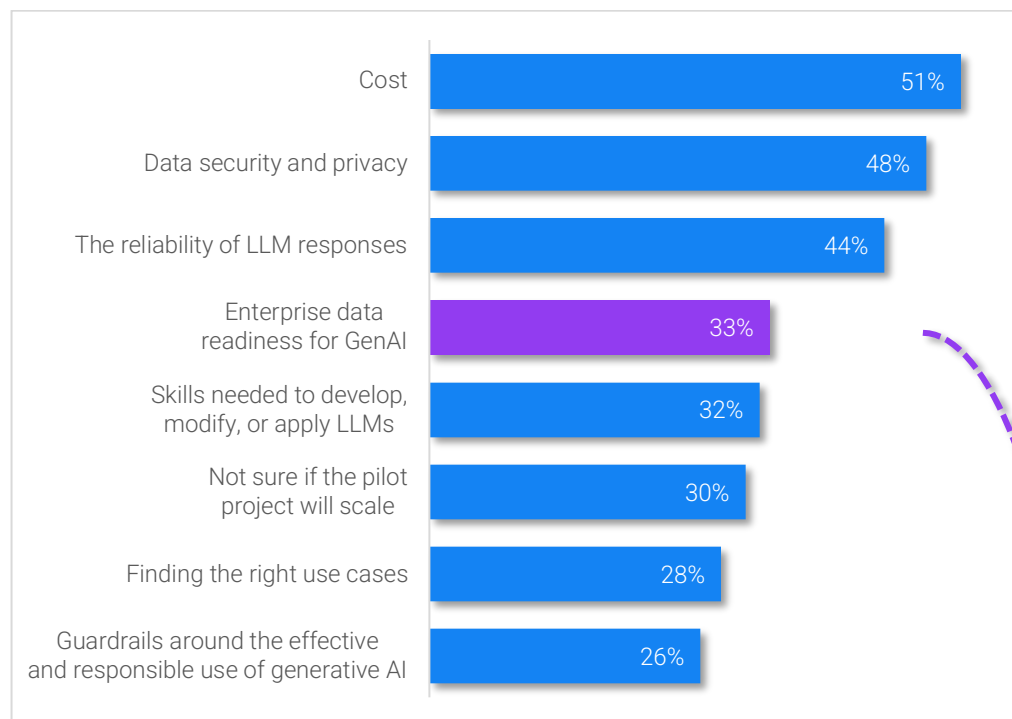


Figure 8: Top challenges in deploying GenAI applications



Figure 9: Enterprise data as an obstacle, by company size

## Top concerns in leveraging enterprise data for GenAI and RAG (all respondents)

Scalability and performance (48%) are the biggest concerns in leveraging enterprise data for GenAI and RAG. Data quality and consistency (46%) present another significant challenge, especially in large organizations where enterprise data is often dispersed across numerous systems and applications, leading to data duplication and inconsistencies.

Real-time data integration and access (46%) is also a critical concern. For example, in customer-facing applications like customer service chatbots, the ability to access accurate, up-to-date customer data in real time is essential for providing relevant and reliable responses.

Additionally, data governance and compliance (44%), security and privacy issues (43%) remain paramount when dealing with sensitive enterprise data.

Notably, these concerns are distinct from the general obstacles faced in GenAI deployment (fig. 9). With the top concerns so close together in percentage, it's clear that addressing only one issue, such as scalability, without ensuring data quality or security, is insufficient. A holistic approach is needed to effectively scale and integrate GenAI projects with enterprise data, ensuring all challenges are addressed.

\*Question allowed more than one answer and as a result, percentages will add up to more than 100%

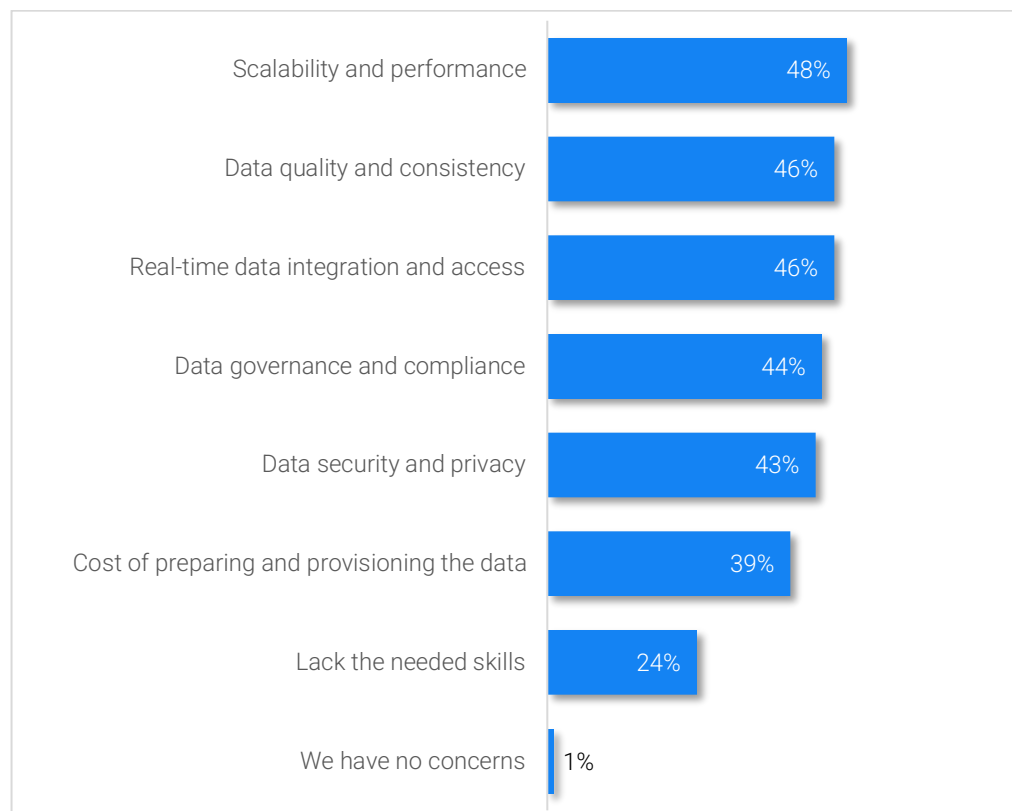


Figure 10: Top concerns in leveraging enterprise data for GenAI and RAG (all respondents)

## Top concerns in leveraging enterprise data for GenAI and RAG (by industry)

While scalability and performance (48%), data quality and consistency (46%), and real-time data integration and access (46%) are top concerns for leveraging enterprise data in GenAI and RAG across all industries, the specific challenges vary by sector.

In the Retail and Telco industries, data quality and consistency clearly emerge as the most significant issues. This can be attributed to the history of mergers and acquisitions in these sectors, leading to data being sourced from disparate systems with varying levels of consistency.

Security and privacy clearly stand out as top concerns for the Financial Services and Healthcare industries. This makes sense given the private nature of the data in these industries.

The ability to scale GenAI applications while maintaining performance standards is a universal challenge.

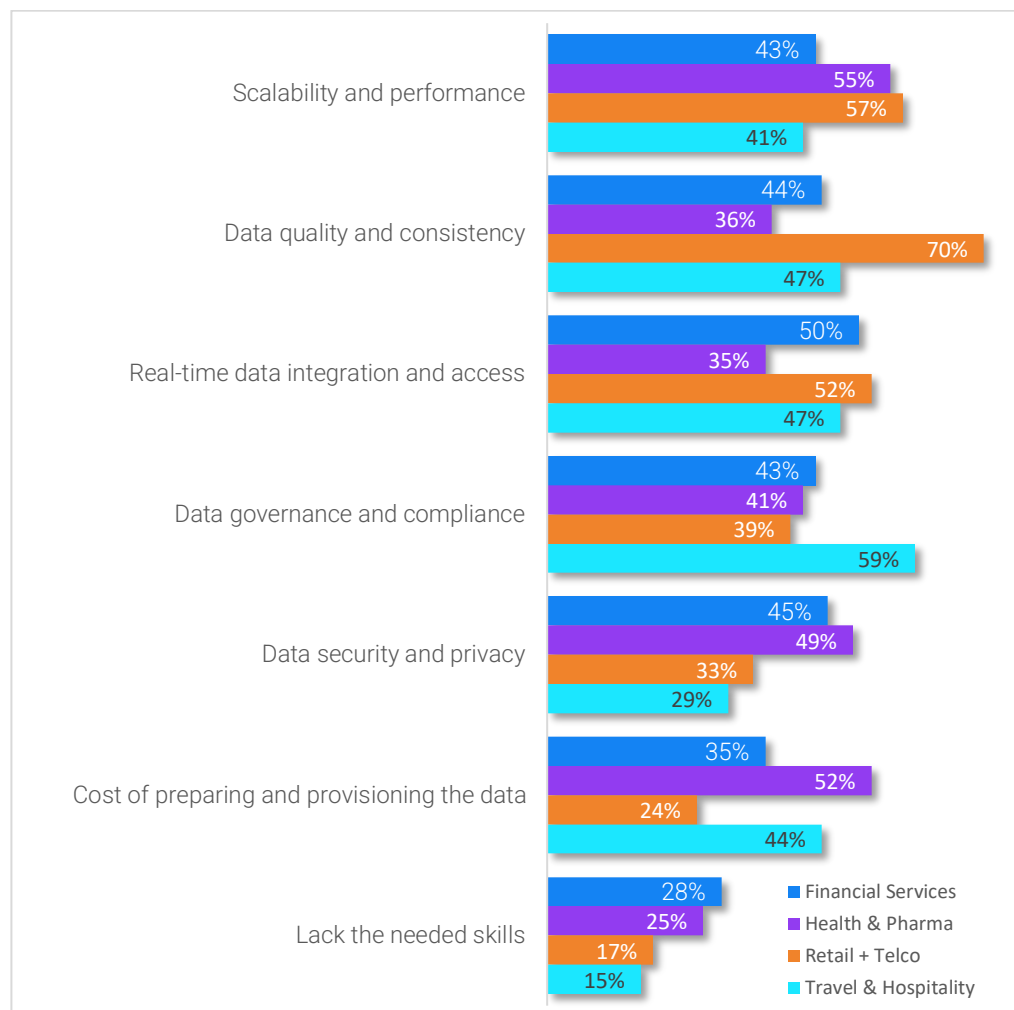
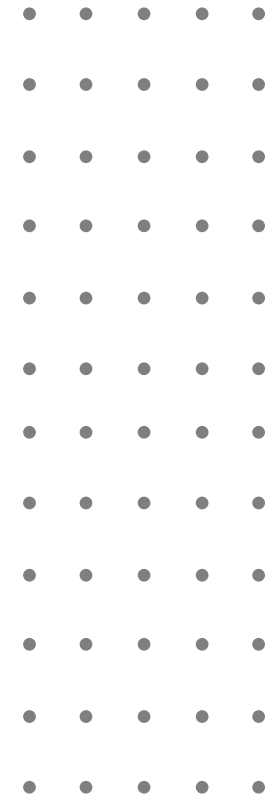


Figure 11: Top concerns in leveraging enterprise data for GenAI and RAG (by industry)

\*Question allowed more than one answer and as a result, percentages will add up to more than 100%



# Data-readiness

## Most valuable internal data sources for customer-facing GenAI

Our findings show that valuable data for customer-facing GenAI applications is spread across multiple sources, creating significant challenges in data integration and governance.

Customer interactions and transactions—covering invoices, payments, products, calls, chats, and emails—are identified as the most valuable internal data source (56%). Close behind is CRM data (55%), which contains critical customer information such as support tickets, promotions, sales, purchase history, and preferences. Marketing data (54%) also plays a key role, including campaign performance and customer engagement metrics.

Interestingly, no single source dominates. Instead, valuable data is dispersed across various departments. This highlights the need for organizations to integrate and harmonize data from multiple internal systems to fully leverage GenAI for customer-facing applications.

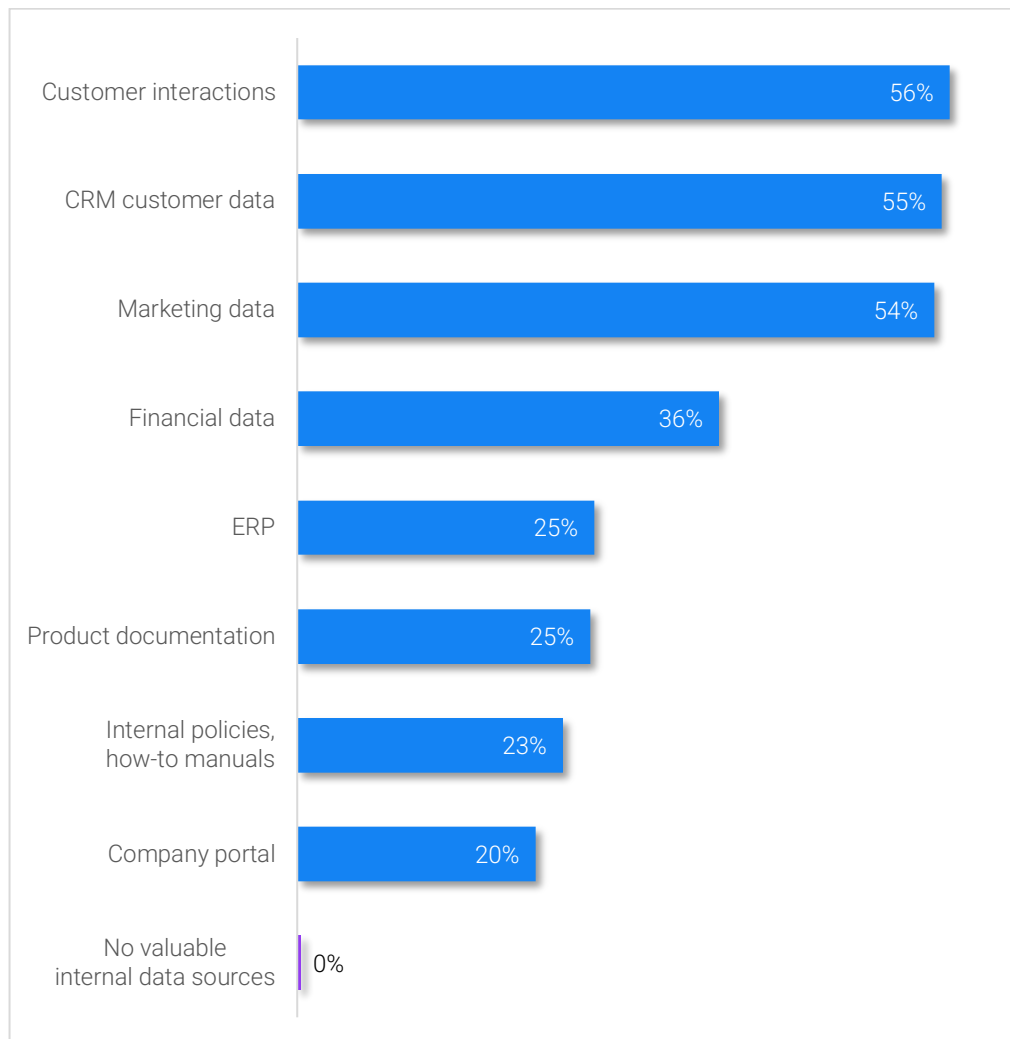


Figure 12: Most valuable internal data sources for customer-facing GenAI

\*Question allowed more than one answer and as a result, percentages will add up to more than 100%

## Top systems housing customer data

The primary systems housing customer data are CRM sales platforms (57%), campaign management tools (53%), and finance systems (53%).

These findings indicate that customer data does not reside in a single location but is distributed across various platforms.

This fragmentation of customer data across multiple systems reflects the complexity of managing and leveraging data in enterprise environments. Just as the value of customer data is derived from various sources, the data itself is stored in a variety of applications. To effectively utilize customer data in GenAI applications, organizations must ensure that customer data is unified from each of these systems, and governed for use in GenAI applications.

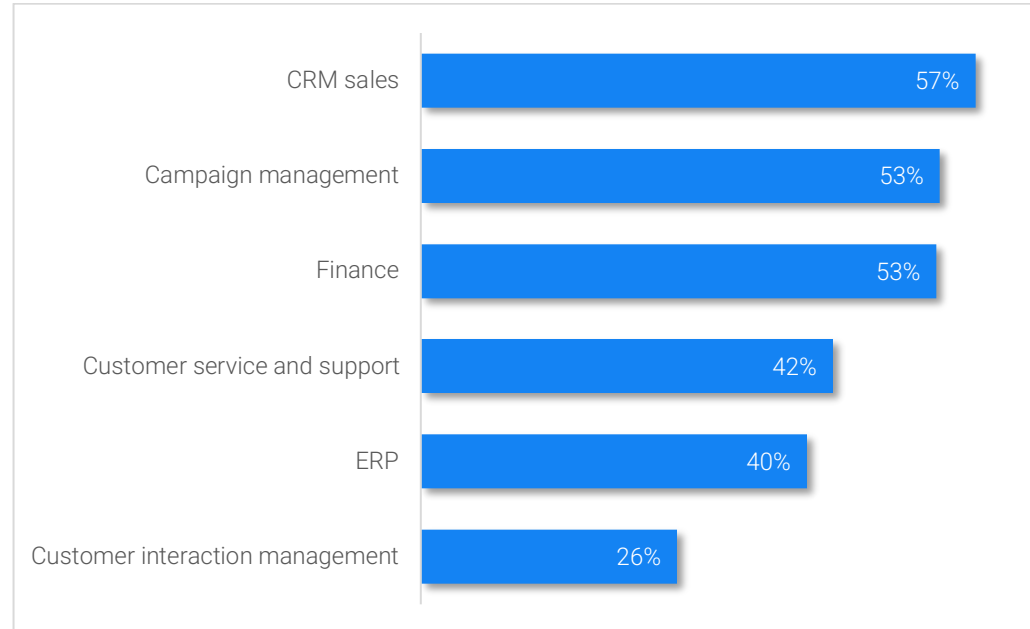


Figure 13: Top systems housing customer data

\*Question allowed more than one answer and as a result, percentages will add up to more than 100%

## Types of internal data used for GenAI and RAG projects

The most utilized internal data sources for GenAI and RAG projects are data from operational systems (54%), data from data lakes and warehouses (45%), vector databases (44%), and graph databases (38%).

These findings highlight the diverse approaches organizations are taking to harness internal enterprise data for their GenAI initiatives. Companies are working with both unstructured data sources, such as vector and graph databases, as well as structured sources like operational systems and data lakes.

The use of multiple data sources underscores the complexity of integrating enterprise data for GenAI and RAG projects. Each source provides unique types of data and offers distinct advantages, but the true value lies in combining them. Relying on just one type of data is insufficient, so organizations must carefully select a data strategy that aligns with their specific needs and goals.

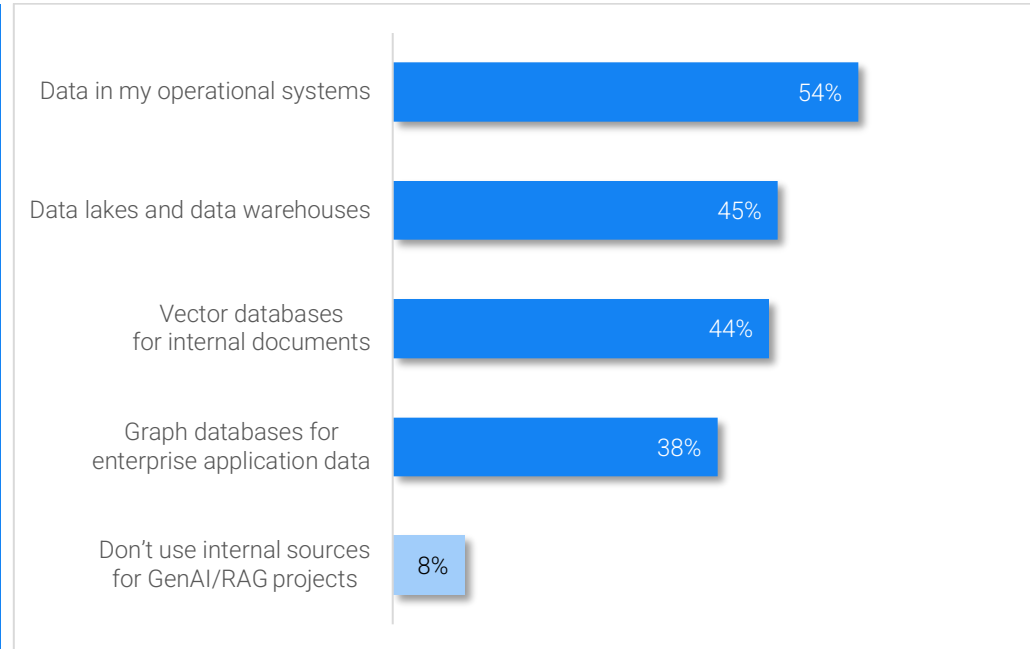
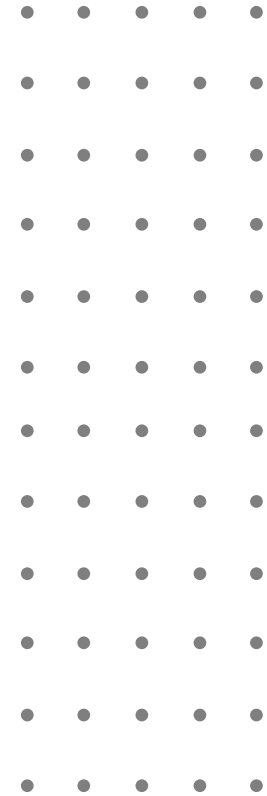


Figure 14: Types of internal data used for GenAI and RAG projects

\*Question allowed more than one answer and as a result, percentages will add up to more than 100%



# Demographics

## Country, target audience, company size, industry, department, and job seniority

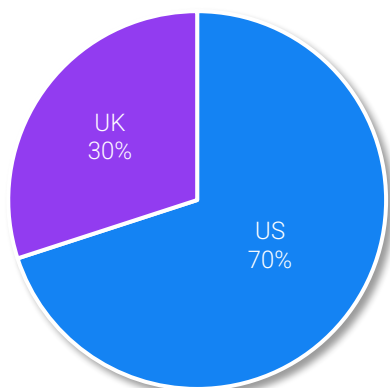


Figure 15: Country

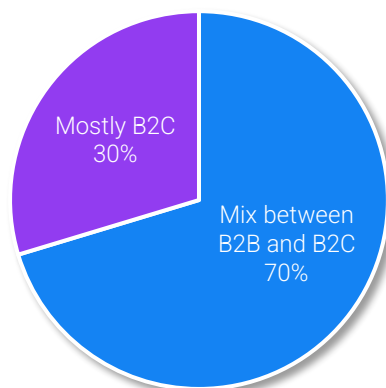


Figure 16: Company's target audience

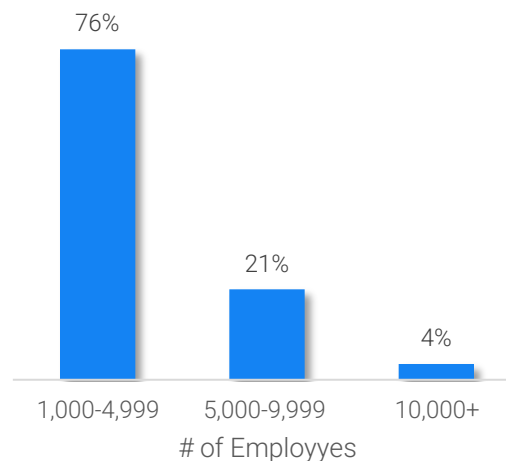


Figure 17: Company size

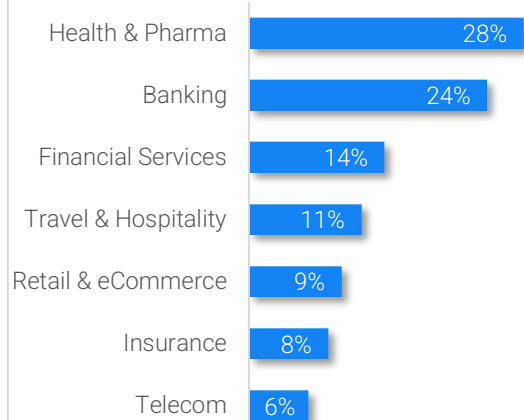


Figure 18: Industry

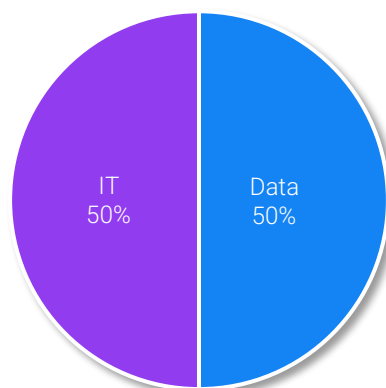


Figure 19: Department

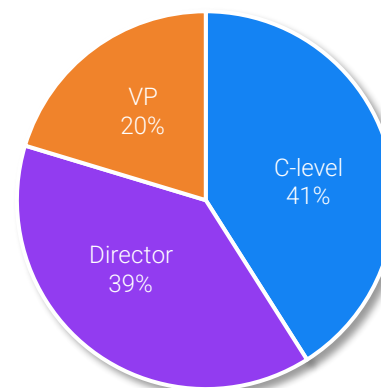


Figure 20: Job seniority

## Personal involvement with GenAI technologies

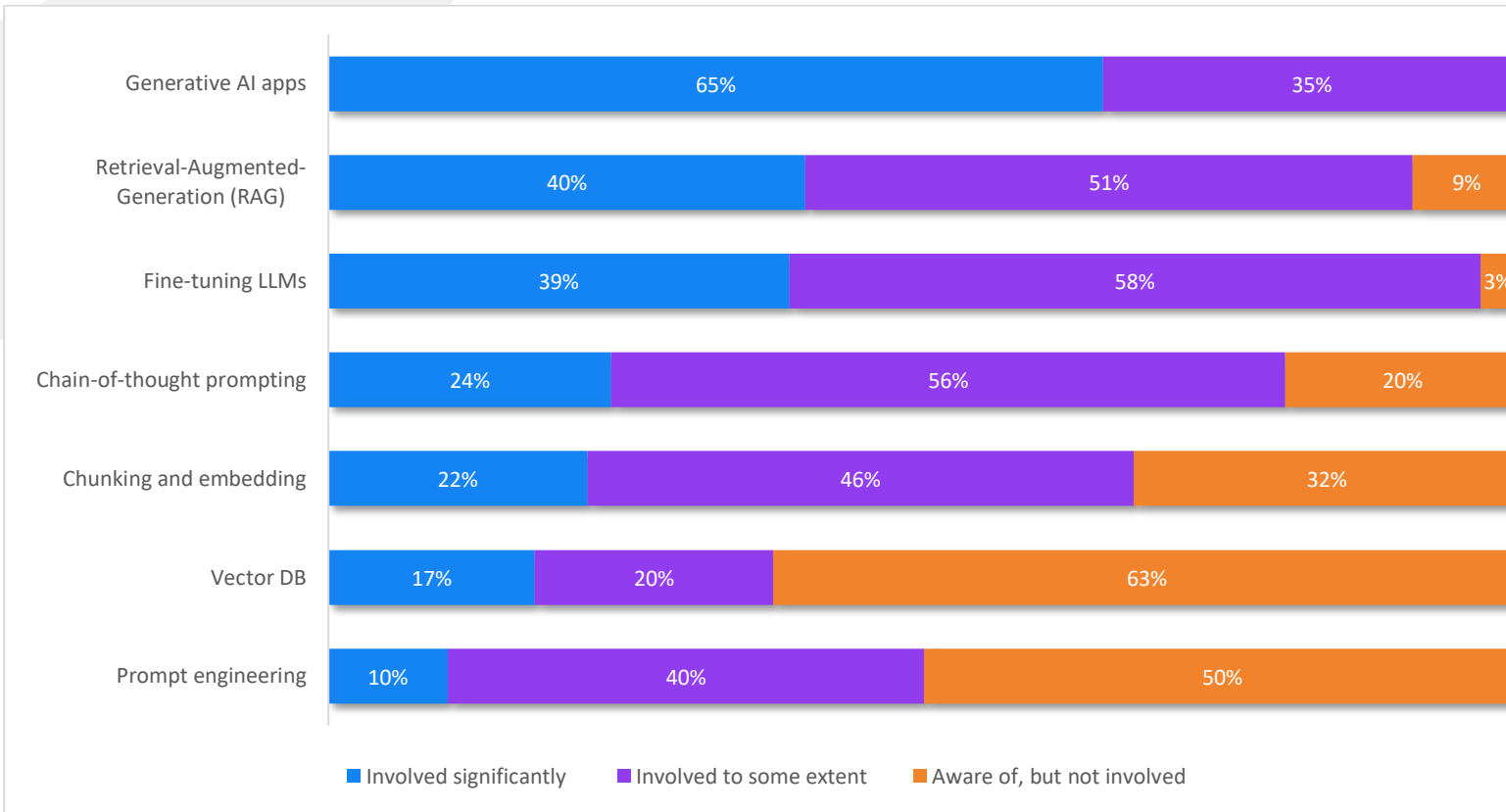


Figure 21: Personal involvement with GenAI technologies

## About K2view

---

In recent years, the basis of data integration platforms have morphed from API, to cloud, through iPaaS , and now, AI and Generative AI. Throughout these evolutions, one platform has proven to be particularly flexible, adaptable and future proof: Data Product Platform by K2view. Why? Because it provisions structured data – via generative data products – that’s always complete, protected, and accessible in real time (in other words, AI-ready).

And, once generated, you can reuse your data products to power your generative AI apps to respond accurately and securely to any query by any user using frameworks like RAG (retrieval-augmented generation), and across a variety of use cases like synthetic data generation, test data management, and data masking,

Data Product Platform deploys in weeks, on-prem or in the cloud, scales linearly, and supports some of the largest enterprises in the world. For all these reasons, and more, Gartner rates K2view a Visionary – testifying to our ongoing commitment to innovation and value delivery.

[Learn more](#)

For more information, please visit us:

