IDC

SPOTLIGHT

Sponsored by: IBM

To succeed in a world that is suddenly "AI everywhere," organizations need the right tools to develop, manage, orchestrate, and optimize AI at scale.

Five Key Orchestration Capabilities to Achieve AI Value at Enterprise Scale

September 2024

Written by: Hayley Sutherland, Research Manager, Conversational AI and Intelligent Knowledge Discovery

Introduction

Generative AI (GenAI) is the new wave of advanced AI. With its large language models (LLMs) that respond to and generate humanlike language, it has great promise for business use. As of February 2024, 53% of organizations that IDC surveyed either had already introduced several GenAI applications/services into production and were focused on expansion or were investing significantly in GenAI, with an established spending plan for training and acquiring GenAI-enhanced software. Another 31% were conducting the initial testing of models and focused proof of concepts (POCs), according to IDC's February 2024 *Future Enterprise Resiliency and Spending Survey, Wave 2.*

AT A GLANCE

KEY STATS

IDC's July 2024 *Global GenAI Technology Trends Survey* found that common barriers to operationalizing GenAI include:

- Lack of developers with the required AI skills (38.2%)
- Inadequate infrastructure performance/availability (36.9%)
- Inability to access required data sets (33.7%)
- Misalignment of use case scope/requirements (31.1%)

However, while GenAI has been grabbing recent headlines, business leaders should not forget about other forms of AI, such as predictive and interpretive AI, which have been in business use for years and have proven value. Strategic, thoughtful use of different forms of AI in combination can multiply the value of newer, advanced technologies such as GenAI. For example, when organizations use intelligent automation, AI-powered search, and predictive AI in conjunction with newer forms of AI, such as GenAI/LLMs, they have the potential to truly unlock the value of their business data, gaining valuable insights and automating repetitive, mundane tasks. In turn, this frees and empowers humans to focus on more complex, creative, and valuable work.

Business leaders across a wide range of industries are experiencing pressure and urgency to deploy AI, particularly GenAI, so that their organizations are not left behind. However, as we saw with other forms of AI, one of the biggest emerging challenges is the gap between the urgency to deploy GenAI and organizational readiness to do so effectively, particularly from a data readiness perspective. For example, in IDC's January 2024 *Future Enterprise Resiliency and Spending Survey, Wave 1,* just 24% of respondents believed that their resources were extremely prepared for GenAI and only a third or less of those organizations could ensure that GenAI models were using high-quality data, understood where personal and corporate IP sensitivities existed in data shared with GenAI models, and could trace what GenAI models were doing with the data being sent through prompts. Common barriers to operationalizing GenAI include a dearth of developers with the required AI skills (38.2%), inadequate infrastructure performance/availability (36.9%), the inability to access

required data sets (33.7%), and a misalignment of use case scope/requirements (31.1%), according to IDC's July 2024 *Global GenAI Technology Trends Survey.* These issues can be significant hurdles to gaining value from AI and expanding the use of new technologies like LLMs beyond an initial pilot or POCs.

Becoming an AI-driven organization — one that uses AI on a companywide and centrally managed scale to gain a competitive advantage — demands considerable investment in time and resources. To gain value from this process and become future-ready, organizations need to close the AI readiness gap. This requires a strategic, organizationwide approach to AI, as well as the right tools. Five key orchestration capabilities that will help organizations close the AI readiness gap, leverage advanced AI capabilities at scale, and translate AI into a business advantage are outlined in the sections that follow.

Capability One: Data Readiness

As discussed previously, data readiness challenges are one of the biggest issues organizations face in using AI to develop useful, actionable insights and predictions. Organizations should develop processes internally, leverage professional services, and/or seek software vendors and products that help automate the following aspects:

- **Data management and classification:** How is data captured, tagged, classified, and stored?
- **Data selection/curation:** Which data is the right data to achieve the desired outcomes?
- **Data quality:** How fresh and accurate is the existing data? Which applications will require "real-time" versus asynchronous data? And how often will it need to be refreshed?
- **Data enrichment:** How can existing data be made more accurate and/or more detailed (e.g., with metadata tagging)?
- **Data transparency:** How does data flow throughout business/information technology (IT) systems in given workflows?
- **Data security:** How is data secured — from ensuring there are no IP or PII data leaks to understanding and implementing access permissions?

Capability Two: Use Case Selection and Expansion

Organizations need to prioritize valuable, feasible, and responsible AI use cases, especially when it comes to newer forms of AI, such as GenAI/LLMs. The goal should not be to completely automate away a human workforce but rather to strategically use AI and automation technologies to augment human work and optimize business workflows. However, many organizations are new to understanding how to select such use cases and align them with different AI product features, how to select and present metrics to show current success and future potential, and how human workers can be upskilled to work with and alongside such tools. To support this capability, organizations should consider developing a center of excellence and working with internal and external experts, AI-centric technology vendors, and other partners to determine which use cases are the best to start with, how to measure them to prove broader business value, and how to expand upon those use cases to use AI at scale throughout the organization. AI product selection should focus on features that facilitate clear, flexible ROI measurement and deployment of proven AI use cases.

Capability Three: Deployment Speed and Flexibility

On a related note, the right AI orchestration platform should not only facilitate but accelerate AI development and deployment so that organizations can achieve faster time to value. Technology vendors with experience and expertise in AI use cases (and that have related data sets) can help customers leverage repeatable aspects and speed development while lowering the risk and cost of false starts. Features supporting this capability can include quick-start elements such as prebuilt modules and AI skills as well as low-code and no-code tools to democratize and speed up AI development and improvement.

Capability Four: Enterprise-Scale Interoperability

Taking advantage of advanced technology at enterprise scale requires enterprise-scale interoperability. In today's world of automation, AI assistants, and other forms of AI, there are three major levels of interoperability to consider:

- **Data interoperability:** The platform must easily connect to existing data, business knowledge, and systems of record. The aforementioned capability one will help organizations determine which data and related sources they should prioritize. It may also be important to consider when different data types, especially structured and unstructured, will need to be used together.
- **Workflow interoperability:** AI systems are increasingly connected to automation systems to automate tasks and workflows. By enabling interoperability and easy integration with these or other key software systems, such as development tools or scheduling systems, AI assistants can help automate tasks and provide "agentic" capabilities, such as autonomously completing workflows.
- **Assistant interoperability:** As AI assistants proliferate across different use cases, applications, and platforms, the ability to orchestrate — that is, understand and manage different assistants, their primary functions, and alignment with user needs — will become increasingly important. AI assistants that can work in conjunction with other assistants and automate this at a higher level based on such alignment models provide even more value for a businesswide approach to AI.

Capability Five: Continuous Improvement

Finally, continuous improvement is another key capability that supports the successful, valuable use of AI on an organizational scale. One of the differentiating values of advanced AI, from predictive machine learning (ML) models to large language models, is its ability to constantly learn and improve from real-world feedback and additional data. Features that support this capability include:

- Feedback mechanisms (e.g., giving thumbs up/down, asking "did this answer your question?")
- Manual and automatic tuning controls
- The ability to leverage new data
- The ability to infuse and leverage new technological innovations to provide greater value over time

Definitions

- **Multimodal:** Multimodal refers to multiple modes or types of data being used simultaneously. For example, multimodal AI systems use a combination of two or more types of data, such as texts and audios, as input to conduct analyses, make predictions, or generate content and insights.
- **GenAI:** Generative AI is a branch of AI that uses unsupervised and semi-supervised ML algorithms to generate new content using previously created content as data (e.g., texts, audios, videos, images, and codes).
- **LLMs:** An example of GenAI, LLMs are language-generation models with vast numbers of parameters. LLMs began at Google Brain in 2017, were initially used to translate words while preserving context, and have since proliferated. In addition, some organizations have developed their own LLMs, or customized models based on OS and commercial versions for business use.
- **Retrieval-augmented generation (RAG):** RAG is an approach to answer generation that leverages the strengths of both search and generative AI technologies. In RAG, a search system first retrieves relevant information (typically via vector search, also known as semantic or similarity search) and then passes it to the LLM in the form of best-fit snippets or documents, which the LLM then summarizes into an answer or insight.
- **Skills:** These are prebuilt, reusable AI capabilities for automating repeatable actions, answering questions, and other tasks.

Benefits

In 2023, 83% of surveyed IT leaders believed that GenAI models that leverage their own business' data would give them a significant advantage over their competitors (source: IDC's *GenAI ARC Survey,* August 2023). Automating IT tasks (46.5%) and automating business processes and workflows (31%) were the top AI use case categories already providing ROI to businesses in late 2023, as well as the top 2 use case categories that organizations expected would bring the greatest ROI over the next 24 months (source: IDC's *Global AI [including GenAI] Buyer Sentiment, Adoption, and Business Value Survey,* October 2023). This shows the value organizations are already seeing from AI assistants and other AI models that can complete development code, triage IT help desk requests, answer customer and employee questions, conduct automated compliance checks, or monitor IT systems for potential threats. However, the same survey revealed that for 71.2% of respondents, it took one to three years — or in some cases, even longer — to achieve ROI on AI initiatives. This gap between deployment and time to value shows the impact of challenges such as data readiness and lack of in-house AI resources.

All of these factors point to the importance of use case–specific AI, that is, templates and quick-start modules that allow organizations to deploy proven AI use cases faster and with a significantly lower risk of failure. By selecting tools that are also easy to use by workers with different roles and skill sets, from experienced developers to non-developer professionals from the lines of business, organizations can facilitate the use and expansion of AI across the company to achieve greater ROI. This allows humans who are subject matter experts, but not coding experts, to infuse key business knowledge into AI assistants and other applications. Organizations that leverage tools and processes to speed up the development of accurate, useful AI systems for proven value use cases can significantly shorten the time to value.

Considering IBM

IBM has been advancing information technology, including AI, for over a century, providing infrastructure, software, and consulting services for organizations around the globe. With watsonx Orchestrate, IBM brings together GenAI and digital automation functionalities to provide capabilities for building personalized AI assistants and using and managing prebuilt IBM assistants for different domains, such as HR, sales, and customer care. watsonx Orchestrate is part of a broader family of IBM AI assistants that includes watsonx Orchestrate, watsonx Code Assistant, and other purpose-built assistants such as watsonx Assistant for Z. Orchestrate is built on watsonx, IBM's proprietary enterprise AI and data platform, which itself consists of three core pillars:

- **watsonx.ai: Train, validate, tune, and deploy foundation and ML models.** watsonx.ai combines the ability to train and deploy LLMs, as well as other forms of AI such as ML to enable organizations to optimize AI usage. watsonx.ai also incorporates the ability to generate synthetic data for model training, which can help organizations overcome data access, quality, or quantity challenges. It also provides visual GUIs and no-code content generation capabilities so non-developers can also contribute.
- **watsonx.data: Scale analytics and AI workloads.** watsonx.data can help organizations address issues related to data integrity, providing a single point of entry for all data sources with a shared metadata layer across clouds. Admins can unify, curate, and prepare data for AI models within watsonx.data, while workload optimization across multiple query engines and storage tiers can help minimize data warehouse costs.
- **watsonx.governance: Implement governance, transparency, and explainability.** watsonx.governance helps with managing risk and bias in models and provides visibility into data and AI workflows, including performance metrics. In addition to full AI model life-cycle governance and risk management, watsonx.governance can also help organizations keep up with and address compliance policies worldwide, including the EU AI Act.

watsonx Orchestrate leverages these three core pillars of watsonx to provide an assistant builder and skill studio intended to accelerate the process of building successful custom AI agents and assistants. The skill studio provides a catalog of AI skills from IBM and third-party platforms that are tuned to help overburdened knowledge workers, specialists, and other employees with everyday tasks. This allows organizations to strategically automate cumbersome, manual tasks with the right preinfused blend of AI technologies and data capabilities. Skills are also prepackaged into use case–specific assistants for functions such as HR, procurement, and sales for the fast deployment of AI assistants that can tackle broader workflows based on repeatable business use cases with high ROI. In addition to code-based tools designed for developers, low-code and no-code building capabilities for business teams help further speed development and enable different (but equally critical) types of organizational experts to deploy useful AI assistants.

Base LLM capabilities can draw from a variety of IBM-developed or third-party models. Embedded IBM AI models allow IBM customers to leverage the right pretested LLM for each use case, while open integrations with third-party models provide organizations with additional flexibility. watsonx Orchestrate LLM capabilities can be integrated into RAG pipelines, allowing them to "ground" in, or retrieve information from, enterprise data sources to generate more accurate, relevant answers and content.

As the name suggests, watsonx Orchestrate also provides capabilities for orchestrating multiple AI assistants across different workflows and functions. It provides a single entry point to invoke the right agent or assistant, depending on what the user is trying to accomplish. This is becoming an increasingly important capability as AI assistants proliferate

across different platforms, applications, and business functions. As companies first experiment with, expand, and eventually look to consolidate the use of AI assistants, the ability to work across and manage multiple domain- and application-specific assistant workflows will help organizations multiply the value of AI and ensure they can continue to leverage existing investments in AI assistants.

Challenges

IBM is facing an increasingly competitive market characterized by rapid innovation and frequent changes. While the company has a strong history of innovation and is providing rapid AI-driven advances with its watsonx platform, it must continue to differentiate its offerings in the areas of AI, data management, and governance to remain competitive. IBM must continually make customers and prospective customers aware of how its proprietary technologies and capabilities are enhancing business outcomes.

Conclusion

Organizations that want to successfully take advantage of AI, particularly GenAI, at enterprise scale need tools that help them quickly achieve and expand value. By providing capabilities to meet challenges, from data readiness requirements to skill gap shortages to selecting, measuring, and rapidly deploying the right AI use cases, IBM watsonx Orchestrate could provide the key capabilities organizations need to achieve enterprisewide AI value.

Organizations that want to successfully take advantage of AI, particularly GenAI, at enterprise scale need tools that help them quickly achieve and expand value.

About the Analyst

Hayley Sutherland, *Research Manager, Conversational AI and Intelligent Knowledge Discovery*

Hayley Sutherland is a research manager for Conversational AI and Intelligent Knowledge Discovery within IDC's Software market research and advisory group. Her core research coverage includes conversational AI and search, with a focus in AI software development tools and techniques for chatbots and digital assistants, speech AI and text AI, machine translation, embedded knowledge graph creation, intelligent knowledge discovery, and affective computing (also known as emotion AI).

MESSAGE FROM THE SPONSOR

IBM watsonx Orchestrate is a low-code generative AI and automation solution that powers employee productivity and customer experience by automating tasks, simplifying complex processes, and ultimately saving time and effort. With an engaging natural language interface and a broad range of prebuilt apps and skills, Orchestrate helps users complete daily tasks across a range of business domains including HR, sales, procurement and customer care. To learn more and access a 30-day free trial, visit the IBM watsonx Orchestrate website.

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
blogs.idc.com
www.idc.com