TDWI Snapshot Series

The State of Data and Operational Readiness for AI

By Donald Farmer and Fern Halper, Ph.D.

Sponsored by:

MongoDB. Capgemini

tdwi

The State of Data and Operational Readiness for AI

By Donald Farmer and Fern Halper, Ph.D.

Table of Contents

Research Methodology

This report examines the current state of AI and how ready organizations are to implement it from a data and operations perspective. TDWI Research has identified five key areas that are critical for AI success: organizational readiness, data readiness, skills and tools readiness, operational readiness, and governance readiness. This report drills into the data side of AI readiness with a focus on generative AI. It examines the challenges organizations are facing in getting ready for AI. Additionally, it provides considerations and best practices for moving forward with data and operational readiness for AI.

For this study, TDWI examined several surveys and assessments that we run throughout the year. Data from this report comes primarily from the 2024-2025 TDWI AI Readiness Assessment.[1] At the time of this report, 240 respondents have completed the assessment, representing a diverse mix of industries and company sizes. Participants come from sectors including telecommunications (12%), financial services (12%), manufacturing (11%), and software (8%). More than half of the organizations surveyed report annual revenues exceeding $500 million. The figures presented in this report are based on the completed responses from this varied and representative sample.

[1] See 2024 *TDWI AI Readiness Assessment Guide*, online at tdwi.org/pages/assessments/adv-all-tdwi-ai-readiness-assessment.aspx.

The Scope and Importance of AI

From a technological standpoint, AI is an umbrella term encompassing a myriad of methodologies and techniques. It leverages advances in mathematics, computer science, computational linguistics, cognitive sciences, and robotics, among others. Popular AI technologies include machine learning, natural language processing, and neural networks, which collectively drive the intelligent capabilities of modern AI systems. TDWI often sees organizations building AI models to predict churn and other customer behavior, identify fraud, determine when maintenance will be needed, recommend products, and predict disease, among other use cases.

Practically speaking, AI can provide tangible benefits such as deeper insights, increased productivity, improved customer service, and greater operational efficiencies that drive cost savings and stronger top-line growth that delivers larger profits. In TDWI research, for instance, we see that organizations implementing more sophisticated analytics, such as AI, are more likely to derive top- or bottom-line benefits from their efforts than others.[2] In other words, there is real, tangible value from AI.

[2] See 2023 *TDWI Best Practices Report: Achieving Success with Modern Analytics*, online at tdwi.org/bpreports.

Recently, generative AI and agentic AI have emerged as significant new technologies, In TDWI research, we see that the majority of respondents to our surveys say that they are at least planning or experimenting with generative AI. In fact, generative AI ranked as the top analytics investment priority in a 2025 survey.[3] There are numerous use cases for generative AI. Some popular examples include chatbots for field and customer support, summarizing documents, and classifying information in documents, as well as using generative AI as a front end for analyzing both structured and unstructured data. There is also some movement by forward-looking organizations to build agentic applications that perform a series of tasks autonomously. This might be an application that helps customers manage their finances and schedule payments, for instance.

These applications can be transformative, but to achieve their benefits, organizations need to understand the problems they want to solve and have a solid data foundation that supports high volumes of diverse data. The scale and complexity of these new models and applications can be challenging. For instance, AI apps in production must deliver consistent performance, ensuring data is available when needed and systems can scale to support massive user loads.

[3] Unpublished 2025 TDWI Data and Analytics survey.

Definitions

GENERATIVE AI. Artificial intelligence systems that are designed to create new outputs, such as images, music, text, or other forms of media, based on the input data they are trained on.

AGENTIC AI. Artificial intelligence systems designed to act autonomously and proactively toward specific goals or objectives, often without needing continuous direct human oversight.

NATURAL LANGUAGE PROCESSING. A field of artificial intelligence that focuses on enabling computers to understand, interpret, and generate human language.

LARGE LANGUAGE MODEL. A specific type of AI model that's been trained on vast amounts of text data. This training enables it to understand context, generate human-like text, summarize documents, answer questions, and even perform tasks like translation.

VECTOR EMBEDDINGS. Generative AI applications often require input data to be transformed into a format that the model can interpret—commonly vector embeddings. These are numerical representations of data that capture its semantic meaning. Efficient handling of these high-dimensional vectors may require a vector database designed to store, retrieve, and manage vector representations of data. They enable quick access to relevant information by comparing vector similarities.

RETRIEVAL-AUGMENTED GENERATION. Retrieval-augmented generation (RAG) frameworks provide real-time data access for generative AI systems. RAG works by retrieving data from external sources (often vectorized databases) and combining retrieved data with user prompts to generate accurate, natural language responses. In other words, RAG helps provide context.

SEMANTIC SEARCH. Semantic search is an AI-powered search technique that understands the meaning and context of a query rather than relying solely on keyword matching. It leverages natural language processing (NLP), machine learning, and vector embeddings to deliver more relevant results by interpreting user intent and relationships between concepts.

ACCURACY. The proportion of correct predictions over total predictions.

HALLUCINATIONS. AI-generated inaccuracies or fabrications.

PROMPT ENGINEERING. Techniques for structuring queries/prompts to improve model outputs.

BIAS. Systematic distortions in data or in model outputs.

There's also the question of dealing with unstructured data at scale. Traditional AI projects often started with structured transactional data, such as relational tables. But generative AI leans heavily on text, images, or other unstructured data sources. That requires a new approach to data ingestion and storage. There are also concerns around data privacy when dealing with text or user-generated content.

From a business standpoint, there is a "twist"—heightened expectations. Generative AI is in the news, so executives and stakeholders often expect immediate breakthroughs. In reality, it still takes time to define the right use cases, procure the right data, and validate outputs.

At TDWI, we see that organizations want to build generative AI applications that utilize their own company data, whether that be structured, semistructured, or unstructured. That means they need to have the right platforms and architectures in place to support the volume, diversity, and compute needed for AI. It includes ensuring data integrity and putting pipeline processes in place to feed generative AI applications data, including transformed data such as vector embeddings (discussed below). It also includes developing the skills and processes and implementing the tools needed to productionalize and operationalize models and applications. This might include frameworks such as retrieval-augmented generation (RAG) to deliver data to generative AI applications in context.

This report drills into the data and operational readiness of organizations for AI, with a particular focus on generative and agentic systems.

The Overall State of AI Readiness

There are numerous interrelated factors that contribute to the current state of AI readiness. As mentioned, in addition to data readiness and operational readiness, these include organizational readiness, skills and tools readiness, and governance readiness. In the 2024-2025 AI readiness assessment, the overall median score currently stands at 56 out of 100, which puts respondents midway through the process of getting ready for AI. Not surprisingly, for all respondents the median data and operational readiness scores are similarly in the middle (see Figure 1). This is not surprising and mirrors past patterns.

For instance, back in 2014 TDWI launched its big data maturity model. At that time, we saw similar kinds of scores for managing big data. Now, organizations have made strides to manage high volumes of data. Additionally, at that time, over 35% of respondents said they would never use the cloud to manage big data. Now, utilizing the cloud is considered a best practice for big data. The point is that adopting new technologies such as AI and generative AI

Dimension	Median Score (Out of 20)
Data readiness	**12.2**
Operational readiness	**11.0**

Figure 1. *Participants' median scores for the data and operational readiness dimensions of the AI Readiness Assessment. Based on 240 respondents.*

takes time and typically follows a maturity curve. With time, more widespread adoption and best practices can emerge as organizations mature.

The Assessment results indicate that on average, companies are currently in the process of putting a strategy in place for AI. In terms of data readiness, these organizations are typically collecting more than just structured data, although they may not be analyzing it yet. This may include internal and external text data such as customer trouble tickets, IT incident reports, or other internal and external documents. It may include machine-generated data and real-time data streams. It may include application clickstream data and application data sources.

They are working to make this data easily accessible. They typically utilize a range of data platforms including a data warehouse or data lake, both on premises and in the cloud, to manage this diverse data. These platforms can support building AI models against structured data. Enterprises are trying to implement an architecture to support data growth, but they are not yet there.

Likewise, organizations midway through the AI readiness maturity journey are starting to think about the need to build applications that utilize AI and put these applications into production. However, companies at this stage don't typically have AI developers in place. They don't necessarily have the Ops teams that they will eventually need to put certain operational processes in place for AI in production. These processes include data pipeline management to feed models with high-quality and up-to-date data, model training and fine-tuning to align outputs with business needs, and MLOps (machine learning operations) to automate model deployment, monitoring, and version control. Many have not yet operationalized any kind of generative AI application using techniques such as RAG.

Additionally, continuous monitoring and feedback loops to help detect model drift, biases, and hallucinations (ensuring the AI remains accurate and trustworthy) are not yet in place during this stage. As they build applications, organizations will also need to put frameworks in place, such as software development frameworks. Again, many organizations are not there yet, having good intentions around data and AI without yet building the necessary governance framework or risk management approach. That can lead to a lot of problems, including model bias or potential compliance issues once models start impacting real decisions.

The following sections drill into the data and operational readiness dimensions in more detail.

The State of Data Readiness for AI

Organizations must have the right data platforms in place to support AI, which often deals with massive volumes of diverse data. AI models can be computationally expensive to build and operate, and this is especially the case with

DATA AND OPERATIONAL READINESS FOR AI: INSIGHTS FROM MONGODB'S BENJAMIN FLAST

Benjamin Flast, director of product management at MongoDB, notes that organizations fall across a wide spectrum of readiness when it comes to adopting generative AI. On the data side, many companies already possess the necessary information and are looking to restructure or reorient it for use with large language models (LLMs). Unlike earlier phases of AI development that required model training expertise, today's landscape allows organizations to work with pre-trained models and focus on how they structure and retrieve data for interaction. This lowers the barrier to entry for data readiness, particularly for organizations already managing semistructured and unstructured data.

Operational readiness presents more significant challenges. Introducing generative AI into production environments requires new forms of monitoring and evaluation. Traditional metrics—such as system uptime or error rates—are no longer sufficient. Organizations now need to assess whether systems are delivering relevant and accurate responses, especially in use cases like retrieval-augmented generation (RAG). This involves testing against curated question-answer pairs, evaluating whether the right documents are retrieved, and adjusting as user queries and expectations evolve. Maintaining quality over time requires real-time feedback mechanisms and the ability to detect shifts in how data is being used.

Flast believes there is a shift in who is building these systems. While data science teams remain involved, there is now broader participation from software and application developers working directly with generative AI models. The most successful efforts tend to involve collaboration between developers and data scientists or AI engineers, especially when designing evaluation frameworks or addressing system reliability.

MongoDB supports these efforts with infrastructure that allows for storage and querying of a wide variety of data types, including documents, vectors, and other file types. MongoDB's support for vector search is particularly relevant for applications that require semantic retrieval. According to Flast, "Whether that is combining vector search with lexical search, whether serving billions of vectors, or storing a diverse set of data in an easy way, MongoDB can support this." This functionality, combined with the flexibility of the document model, allows teams to develop and iterate without needing to rework the database as requirements evolve. MongoDB also includes core governance capabilities—such as role-based access controls, query logging, and permissions management.

generative AI. It is critical for an organization to have a solid data foundation to support its AI efforts.

The assessment findings paint a picture of organizations at various stages of data readiness. For instance, the assessment shows 39% of respondents are now managing newer data formats including text and machine-generated content: a significant shift in data collection practices (Figure 2). We have also seen this in other TDWI research. This data foundation needs to provide a unified view of

Which statement BEST describes the kinds of data your organization collects and manages as part of its analytics and AI efforts?

None	12%
Structured data (i.e., tables, records)	16%
Structured data as well as demographic data such as age, location, etc.	16%
All of the above plus semistructured data (XML and similar)	17%
All of the above plus 1-2 of the following: internally generated text data (e.g., emails, call interaction notes, survey verbatim), social media data (blogs, tweets), machine-generated data, geospatial data, real-time event data, audio, video, weblogs, clickstreams, scientific data, demographic data	22%
All of the above, plus 3 or more of the following: internally generated text data (e.g., emails, call interaction notes, survey verbatim), social media data (blogs, tweets), machine-generated data, geospatial data, real-time event data, audio, video, weblogs, clickstreams, scientific data, demographic data	16%

Figure 2. *Based on 240 respondents.*

both structured and unstructured data to power effective AI applications—for example, a customer service chatbot may require access to structured service catalogs alongside unstructured customer interaction histories. To utilize this data effectively, organizations will need to scale their underlying data platform. Yet just 41% of respondents feel confident they possess sufficient compute power to develop and train AI models (Figure 3). Traditional databases are being pushed to their limits as organizations attempt to process and analyze exponentially growing data volumes.

At the same time, accessibility and understandability of data also remain critical challenges, with only about 40% of respondents reporting systems in place that ensure data is easily accessible and can be integrated from various sources (Figure 4). This creates significant barriers to AI adoption and effectiveness. Additionally, organizations will

My organization possesses the advanced analytics capabilities and computational resources necessary to develop, train, and deploy AI models efficiently

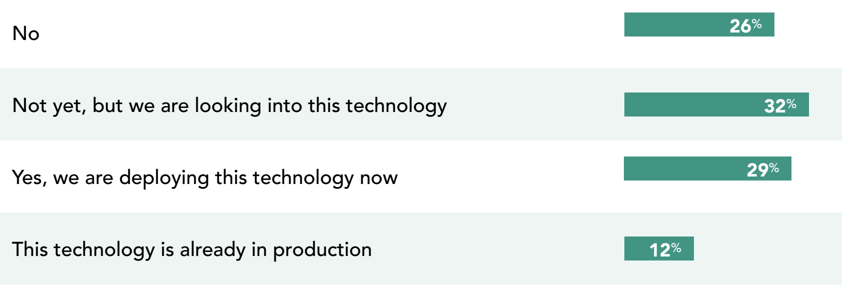

Figure 3. *Based on 240 respondents.*

My organization has systems in place to ensure that data is easily accessible and can be integrated from diverse sources, including internal and external data sets, for analytics and AI applications

Figure 4. *Based on 240 respondents.*

need to address not just technical accessibility but also semantic understanding of data, particularly for generative AI, enabling AI systems to comprehend the meaning and relationships within information rather than simply processing raw values.

Real-time contextual data processing capabilities represent the frontier of AI data infrastructure, with slightly less than half (44%, not shown) of organizations believing they can effectively orchestrate and monitor multiple data pipelines. New approaches to data representation, particularly vector embeddings, are becoming central to AI functionality, yet 38% of organizations have no vector database capabilities, while 30% are still in the investigation phase (Figure 5).

Considerations and Best Practices for Data Readiness

While some respondents are advancing with sophisticated data architectures and capabilities, many still struggle with fundamental data integration and management challenges. The emergence of technologies such as vector databases adds another layer of complexity to the data readiness journey.

My organization makes use of newer technologies such as vector databases to store vector embeddings for use with AI models

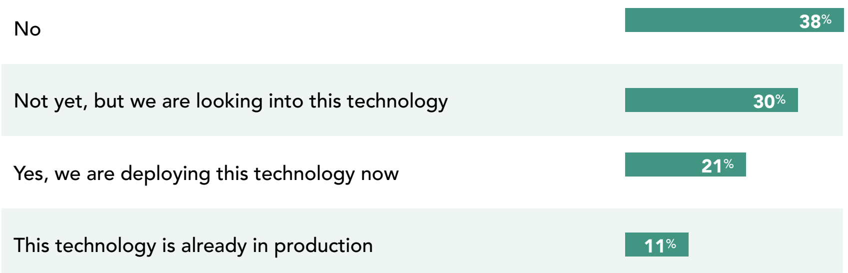

Figure 5. *Based on 240 respondents.*

The assessment data suggests that organizations need to focus on three key areas to improve their data readiness. These include modernizing the data infrastructure to hold diverse data types, building robust data integration capabilities, and developing scalable architectures that can support AI workloads. This is particularly crucial as organizations move toward more sophisticated AI applications, especially generative AI, which requires not only large volumes of data but also sophisticated data processing and storage capabilities.

The gap between ambition and capability remains significant; while many organizations aspire to implement AI solutions, their data infrastructure may not yet be ready to support these initiatives effectively. This suggests that organizations should consider a staged approach to AI implementation, ensuring their data foundation is solid before attempting more advanced AI applications.

The staged approach is needed because there's a difference between a quick pilot project for AI and scaling it across the enterprise. With the volume and variety of data involved, the architecture needs to be robust, unifying structured and unstructured data and enabling near real-time capabilities, such as those needed for retrieval-augmented generation. In a staged approach, organizations address these foundational requirements step by step. (If business users push for advanced generative AI systems without the data foundation, there is a risk of overextending the team and the infrastructure.) This approach involves:

DATA INTEGRATION AND ACCESSIBILITY. Unify disparate data sources into a cohesive environment and improve data discoverability. With so many data types in play, you need to know what data you have and where it resides.

INFRASTRUCTURE AND PIPELINE ORCHESTRATION. Deploy or improve cloud-based orchestration tools, focusing on reliability and scalability. After establishing what data needs to be centralized or integrated, you can design the right scaling mechanisms. Only once you can confirm that data integration and orchestration work reliably should you

move to the next stage of implementing AI frameworks. This way, your environment grows organically and stably, rather than suddenly dumping challenging AI workloads on an ill-prepared system.

EXPANSION TO ADVANCED AI. Once you have the data platform stabilized, you can gradually introduce advanced AI components such as vector databases, large-scale training clusters, or inference pipelines. Generative AI is very demanding; not only does it need extensive compute resources for training and inference, it also benefits tremendously from high-quality, contextually relevant data. In a staged approach, once you have the foundation laid (clean, well-integrated data, plus the ability to handle unstructured data types) then you could add vector search capabilities. If you do it too early, you may be wasting resources on an incomplete solution.

The State of Operational Readiness for AI

With AI, and generative and agentic AI, the ability to build and deploy applications has become even more important. This involves the skills for building applications and putting them into production as well as the tools needed to operationalize AI. Only 11% of respondents to the assessment currently have generative AI models in production (Figure 6). Success in AI implementation depends not just on acquiring technical skills but on creating an integrated ecosystem where technical expertise, business knowledge, and operational capabilities work in harmony.

People in my company are already experimenting with generative AI

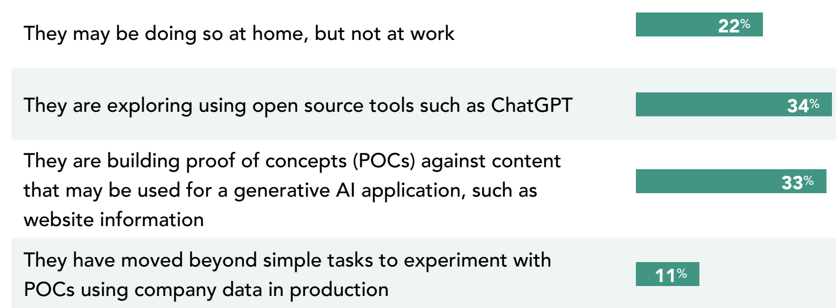

Figure 6. *Based on 240 respondents.*

Skills Readiness

Developing AI applications requires a specialized skill set that extends beyond traditional software development. Teams need proficiency in the entire system development life cycle, coupled with a deep understanding of AI model behavior, including concepts such as metrics of accuracy and latency and model limitations.

The current assessment reveals significant gaps in organizational readiness, with only 56% of respondents reporting adequate data science capabilities. The collaboration gap is even more concerning: only 12% of organizations have data scientists effectively working with business stakeholders, while just 27% have dedicated data engineers in place (not shown). This skills deficit represents a critical bottleneck as organizations attempt to move from AI experimentation to production-ready applications.

EXPERT PERSPECTIVE: MARK OOST, VP, AI AND GENERATIVE AI GROUP OFFER LEADER, CAPGEMINI

AI is now pervasive across industries, with growing budgets and accelerating deployment momentum. According to Mark Oost, Capgemini's VP of AI and generative AI group offer leader, client readiness for AI has noticeably evolved. "We're well past the days of exploring whether AI is relevant," he explains. "Now, every organization sees the value. The question is no longer 'if' but 'how fast and how effectively' they can scale it."

What's emerging is a widening gap between early adopters and those still struggling with foundational capabilities. Organizations that invested early in traditional AI are far more equipped to operationalize generative AI, while others fall further behind. One of the key shifts he sees is in the kind of data organizations prioritize. It's no longer just traditional, structured data—it's operational, real-time, and increasingly unstructured data powering generative and agentic AI systems. In use cases like retrieval-augmented generation (RAG), immediate data access becomes non-negotiable. "If you're serving millions of customers via a generative AI-powered interface, and your app queries a vector database every time, that data has to be ready instantly. Accessing it from cold storage like S3 three million times a day isn't feasible."

To address these challenges, Oost stresses the need for a robust data foundation characterized by availability, traceability, transparency, and accountability. AI today isn't just about building models—it's about integrating them into real-time applications. That integration demands rapid data movement, secure storage, and well-governed flows, particularly in multi-agent environments where data passed between agents must be tightly controlled. Feedback loops and human oversight are also critical for course correction and quality assurance.

Capgemini's RAISE (Reliable AI Solution Engineering) platform was developed to meet these demands. It accelerates deployment of generative AI applications while offering operational tools to monitor, govern, and optimize them in production. MongoDB plays a foundational role in this framework, serving as a fast and flexible vector and document database. "We needed a data layer that could support real-time interaction, ingest unstructured formats like PDFs and PowerPoints, and maintain auditability across the board," Oost says.

He also highlights Capgemini's CPGA (Cloud Private GenAI Agents) solution, which allows organizations to develop agents in the cloud while keeping sensitive data secure—often outside of hyperscaler environments. Together with MongoDB, Capgemini is enabling enterprises to move from isolated proofs of concept to governed, production-grade AI systems.

Tools and Technical Readiness

Organizations are increasingly looking to leverage their proprietary data assets in AI applications, particularly through techniques such as retrieval-augmented generation (RAG) that enhance foundation models with company-specific information. Despite this interest, the tool landscape remains underdeveloped, with only 32% of organizations reporting effective use of AI development tools and frameworks (Figure 7). This suggests that even teams with development resources may lack specializations needed to build robust AI systems.

Real-time data processing capabilities are also becoming essential for AI applications that require immediate contextual understanding and response. In addition to

My organization has access to and effectively utilizes a range of AI development tools, including machine learning frameworks and libraries (e.g., TensorFlow, PyTorch, AI APIs) that are essential for building and training AI models

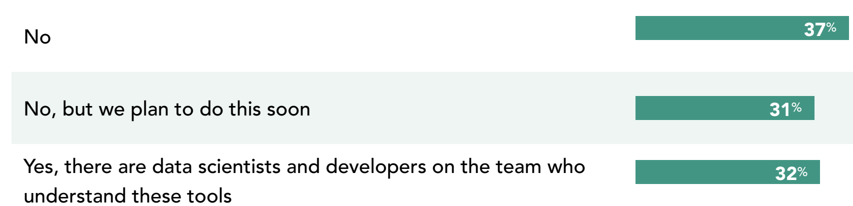

Figure 7. *Based on 240 respondents.*

RAG, semantic search represents a key component of this ecosystem, allowing applications to understand user intent (via business semantics) rather than simply matching keywords. As user expectations shift toward almost instantaneous, contextually aware interactions, organizations will find it necessary to develop infrastructure that supports not just data retrieval but meaningful interpretation of information in real-time, especially for customer-facing applications where response latency directly impacts user satisfaction.

Ensuring accuracy in generative AI applications presents unique challenges compared to traditional software systems. Organizations must implement strategies to mitigate hallucinations—AI-generated content that appears plausible but is factually incorrect. High-quality vector embeddings, such as domain-specific embedding models, are important here. Additionally, techniques such as response verification and source attribution can improve the user acceptance of AI results. To improve accuracy, organizations may consider technical approaches such as re-ranking—where the system generates multiple candidate responses, then evaluates them through a secondary model that scores each option based on factual accuracy, relevance, and alignment with verified sources.

This process separates the generation aspect of generative AI from critical evaluation, allowing the system to first produce fluent, creative content, then filter it through a quality control mechanism that prioritizes accuracy, reducing the risk of hallucinations.

Operational Practices Readiness

Beyond technical approaches, operational practices traditionally associated with MLOps remain essential but require adaptation for generative AI workflows. TDWI research indicates that while many organizations are focused on developing AI capabilities, far fewer are prepared to effectively operationalize and maintain these systems, with MLOps practices still in the early stages of adoption across industries. In this assessment, only about a quarter of the respondents have teams to address AI models in production (Figure 8).

Assessment findings suggest that while organizations are making progress in building AI capabilities, there remains a significant gap between having basic AI skills and achieving the operational maturity needed for successful AI implementation. The data indicates that organizations need to focus not only on hiring data scientists but on

My organization employs MLOps to deal with AI models in production

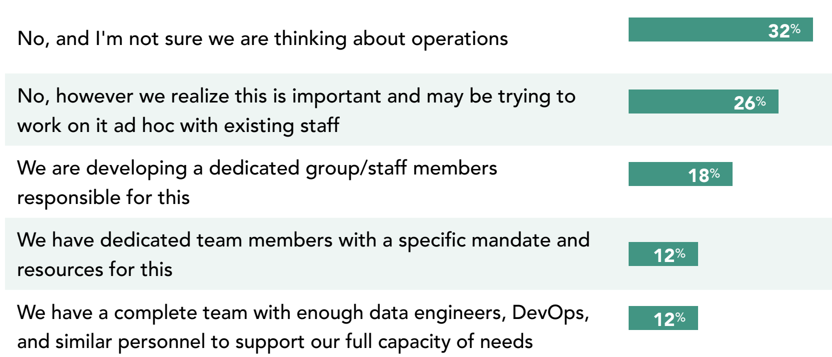

Figure 8. *Based on 240 respondents.*

building comprehensive teams that include data engineers, MLOps specialists, and developers while simultaneously developing training programs to upskill existing staff and building out AI apps for business users.

Considerations and Best Practices for Operational Readiness

In TDWI surveys, only a small portion of respondents have generative AI models in production. As we have noted, this suggests there's a gap between ambition and operational capability.

A big AI risk is discovering "late" that the training data was incomplete or out of date because of poor data pipeline management or data governance. In such cases, the organization doesn't find out until the model starts producing bad recommendations in a live environment. Proper operational readiness means setting up real-time checks on accuracy and usability of the AI. It means setting up robust monitoring, so you don't end up with inaccurate responses that directly impact users. With generative AI, the risk can be even higher because those models can "hallucinate."

To address these issues, it will be important to address issues of governance and compliance, MLOps and deployment pipelines, and new tools and technologies.

GOVERNANCE AND COMPLIANCE. From a governance lens, organizations need clear policies about acceptable data sources. Strong governance frameworks are needed to enforce data privacy, compliance, and usage rights while maintaining role-based access controls and clear data lineage tracking. If you don't have a handle on data lineage (who generated the data, where it's stored, how it's processed) then deploying AI models on top of that data is risky. Any untracked data or poor data quality can cause major trust issues once the model goes live.

Additionally, if pulling in user-generated content from external systems, you must confirm that it's properly vetted for privacy considerations. It will be important to ensure that data privacy laws are respected. If the model calls on private user data, you should have encryption at rest and in transit, plus role-based access controls. The moment you scale generative AI across an entire enterprise, any compliance gap will be amplified.

Cross-functional collaboration is also critical—data engineers, machine learning specialists, software developers, and business users must work together to ensure AI models align with real-world needs and deliver reliable, business-ready outputs.

MLOPS AND PLATFORM ENGINEERING. Teams also need to think about the practical implementation of MLOps which combines best practices from DevOps (software development and operations) with data science and machine learning to streamline the deployment pipeline processes that are part of the AI life cycle. As mentioned, only about 44% of respondents to this assessment believe they have effective orchestration for multiple data pipelines. Organizations must establish robust data pipelines to ensure high-quality, up-to-date data flows

into AI models. In a best-case scenario, you will have an automated system where data is ingested, validated, and fed into model training pipelines. Then, in a separate pipeline you may deploy the model into production, complete with continuous monitoring, so if performance dips or the model's error rate rises, it triggers an alert.

Additionally, it will be important to think about separating and managing the computing resources, data, and operations used for generative AI tasks—such as model training, fine-tuning, or inference—from other workloads in a system or platform. This can help with performance, security, and cost control. In larger organizations, this generative AI workload isolation might fall to DevOps or platform engineering teams. In smaller organizations, it might fall to MLOps or another group.

NEW TOOLS AND TECHNOLOGIES. New tools and technologies are also important. As mentioned earlier, generative AI can require specialized vector storage to do semantic search. That's a newer technology for many. Organizations need to ensure they integrate such solutions thoughtfully into their data architecture; avoid rushing in without a plan.

An important area here is RAG. In RAG, the model uses your internal databases to fetch context, so the AI's answers reference real company data rather than random text from web-scale training. Governance must confirm that the data feeding those retrieval queries is curated, correct, and compliant. RAG can help reduce hallucinations by grounding the model's responses in actual company data.

But it's important to design a feedback loop, so that if the model references data that isn't in the knowledge base, you have a fallback or at least disclaimers. It can also be useful to implement a re-ranking technique so that the system can weed out incorrect results before showing them to the user.

Of course, we cannot overlook skills. If you do not have a team that understands distributed systems and pipeline orchestration, you'll hit a wall. We strongly suggest building cross-functional teams: data engineers, machine learning engineers, software developers, and governance experts collaborating early and often.

This may sound like a lot of moving parts, but this is why the emphasis on "operational maturity" is crucial. Operationalizing AI is not just data scientists building something in isolation—everyone from data engineering to development and governance must align. By adopting a strategic, step-by-step approach, organizations can transition from AI experimentation to full-scale, governed, and high-performing generative AI applications.

Conclusion

Organizations recognize the transformative potential of AI, particularly generative and agentic AI, but success depends on data readiness and operational readiness. This report has highlighted that organizations are making progress but are not there yet in these two dimensions of AI readiness.

Data readiness will require modern infrastructure capable of handling structured and unstructured data, ensuring accessibility, governance, and scalability for AI workloads. Without a strong data foundation, AI applications risk running on incomplete or low-quality data, leading to unreliable results.

Operationalizing AI will require skilled teams, scalable deployment strategies, and governance frameworks to ensure compliance and reliability. Many organizations struggle with MLOps adoption, data pipeline management, and real-time monitoring, limiting their ability to scale AI applications. For generative AI, vector search is becoming critical for retrieving semantically relevant information, improving AI accuracy, and enabling better decision-making. Investing in vector databases and optimizing retrieval techniques such as RAG can help ensure that AI outputs remain grounded in contextually relevant, high-quality data, reducing hallucinations and misinformation.

About the Authors

Donald Farmer is a data and analytics strategist with over 35 years of experience in the field. As principal of TreeHive Strategy and a research fellow at TDWI, he advises global clients on innovation, analytics, and AI. Donald's career includes leadership roles at Microsoft and Qlik, where he pioneered significant product designs. A prolific author and speaker, Donald writes regularly about data management, AI, and innovation. His expertise spans from advising tech giants to guiding start-ups, investors, and government agencies around the world and he has invested in new technologies with teams from Reykjavik to Manila. Donald's innovative spirit extends beyond technology; he and his artist wife, Alison, reside in a unique woodland home near Seattle, with one of the world's most beautiful treehouses. With his ability to deliver transformative insights, Donald continues to shape the future of data analytics and innovation across diverse industries and cultures.

Donald's website is www.treehivestrategy.com.

You can email him at donald.farmer@treehivestrategy.com or connect with him at linkedin.com/in/donalddotfarmer.

Fern Halper, Ph.D., is vice president and senior director of TDWI Research for advanced analytics. She is well known in the analytics community, having been published hundreds of times on data mining and information technology over the past 20 years. Halper is also coauthor of several Dummies books on cloud computing and big data. She focuses on advanced analytics, including predictive analytics, machine learning, AI, cognitive computing, and big data analytics approaches. She has been a partner at industry analyst firm Hurwitz & Associates and a lead data analyst for Bell Labs. She has taught at both Colgate University and Bentley University. Her Ph.D. is from Texas A&M University.

You can reach her by email (fhalper@tdwi.org) and on LinkedIn (linkedin.com/in/fbhalper).

About TDWI Research

TDWI Research provides industry-leading research and advice for data and analytics professionals worldwide. TDWI Research focuses on modern data management, analytics, and data science approaches and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of business and technical challenges surrounding the deployment and use of data and analytics. TDWI Research offers in-depth research reports, commentary, assessments, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

Provided by MongoDB

Navigating the AI Revolution: Adaptation, Adaptation, Adaptation

In 1999, then-President of Microsoft Steve Ballmer gave a famously high-energy speech in which he said that the "key to industry transformation, the key to success is developers, developers, developers, developers, developers, developers, developers, developers, developers, developers, developers, developers, developers, developers! Yes!"

A similar mantra could apply when discussing how to succeed with AI: adaptation, adaptation, adaptation! Artificial intelligence has already begun to transform how we work and live, and the changes AI is bringing to the world will only accelerate. Indeed, the rise of AI has been compared to the Industrial Revolution,[1] and Anthropic's Dario Amodei has called the AI future "a thing of transcendent beauty."[2]

Every organization, therefore, has to adapt to AI-driven changes in how tasks are accomplished, what work looks like, and the market at large. Businesses rely ever more heavily on software to run and execute their strategies, and to keep up with competitors their processes and products must deliver what end-users increasingly expect from AI: speed, ease of use, and personalization.

To do so—and to do so right—requires having a number of fundamental elements in place: the right tech stack, the right software foundation, and the right execution. All of which MongoDB helps organizations of all shapes and sizes with.

[1] For example, Our World in Data's industrial revolution patents chart and AI research team affiliations chart exhibit very similar curves.
[2] https://darioamodei.com/machines-of-loving-grace#taking-stock

The Road Ahead

As the TDWI report makes clear, organizations recognize the potential AI holds, but there's much work to be done. Indeed, only a small percentage—11%—of survey respondents indicated that they had AI applications in production. Success with AI depends on both "data readiness" and "operational readiness."

Data readiness challenges highlighted in the report include managing diverse data types, ensuring accessibility, and providing sufficient compute power. For example, 39% of companies surveyed manage newer data formats, and 41% feel they have enough compute.

Operational readiness, meanwhile, covers things like skill gaps and machine learning operations (MLOps) processes. TDWI's research about the latter "indicates that while many organizations are focused on developing AI capabilities, far fewer are prepared to effectively operationalize and maintain these systems, with MLOps practices still in the early stages of adoption across industries."

How can organizations overcome these gaps?

The report recommends a staged approach: organizations should focus on modernizing their data infrastructure, building robust integrations, and developing scalable architectures. Operational readiness requires skilling (or re-skilling) teams, devising and scaling AI deployment strategies, and instituting governance frameworks to mitigate risks and to ensure AI output accuracy.

A Paradigm Shift

The report also shows how much AI has changed the very definition of software, and how software is developed and managed. Specifically, AI applications are continuously adapting, learning in real-time, and responding to end-user behavior; they can also autonomously make decisions and execute tasks.

All of which depends on having a solid, flexible software foundation, another thing the report highlights:

"As user expectations shift toward almost instantaneous, contextually aware interactions, organizations will find it necessary to develop infrastructure that supports not just data retrieval but meaningful interpretation of information in real-time, especially for customer-facing applications where response latency directly impacts user satisfaction."

Because the agility and adaptability of software is intrinsically linked to the data infrastructure upon which it's built, rigid legacy systems cannot keep pace with the demands of AI-driven change. So modern database solutions like MongoDB, built with change in mind, are an essential part of a successful AI technology stack.

Keeping Up with Accelerating Change

The software stack can be said to comprise three layers: at the "top," the interface or user experience layer; then the business logic layer; and a data foundation at the bottom.

With AI, the same layers are there, but they've evolved: unlike traditional software applications, AI applications are dynamic. Because AI-enriched software can reason and learn, the demands placed on the software stack have changed. For example, AI-powered user experiences include natural conversational interfaces, augmented reality, and experiences that anticipate user needs by learning from other interactions (and from a variety of data). In contrast, traditional software is largely static: it requires inputs or events to execute tasks, and its logic is limited by pre-defined rules.

The database underpinning AI software must therefore be flexible and adaptable, and able to handle all types of data; it must enable high-quality data retrieval; it must respond instantly to new information; and it has to deliver the core requirements of all data solutions: security, resilience, scalability, and performance.

After all, to take actions and to generate trustworthy, reliable responses, AI-powered software needs access to up-to-date, context-rich data. So without the right (read: adaptable) data foundation in place, even the most robust AI strategy will fail. As Figure 1 illustrates, this sort of adaptability is essential for survival.

Figure 1. *The frequency of change across technology eras.*

Keeping up with AI can feel head-spinning, both because of the many players in the space (the number of AI start-ups has jumped sharply since 2022, when ChatGPT was first released[3]), and because of the accelerating pace of AI capabilities, which Nvidia's Jensen Huang has likened to "a hyper Moore's law."[4] Organizations that want to stay ahead must evolve faster than ever before.

MongoDB was built for this kind of change.

The company's founders understood that relational databases block organizations' ability to scale and innovate. So MongoDB was born to enable continuous iteration, eliminate complexity, and to scale effortlessly. Since then, MongoDB has evolved into a fully integrated, AI-ready database platform, ensuring organizations can consolidate workloads, reduce complexity, and adapt quickly. By reducing the friction associated with traditional databases—such as technical debt that drains engineering resources, budgets wasted on

[3] https://ourworldindata.org/grapher/newly-funded-artificial-intelligence-companies
[4] https://www.youtube.com/watch?v=hw7EnjC68Fw

inefficiency, and rigid infrastructures piled with patchwork fixes—MongoDB frees organizations to transform AI ideas into powerful, dynamic, trustworthy[5] software that delivers an immediate impact.

Execution, Execution, Execution

But AI success requires more than just the right technology: expert execution is critical.

Put another way, the difference between success and failure when adapting to any paradigm shift isn't just having the right tools; it's knowing how to wield those tools. So while others experiment, MongoDB has been delivering real-world successes, helping organizations modernize their architectures for the AI era, and building AI applications with speed and confidence.

For example, MongoDB teamed up with the Swiss bank Lombard Odier to modernize its banking tech systems. The MongoDB-Lombard Odier modernization collaboration led to stunning results: code was migrated 50 to 60 times faster than previous migrations, and repetitive tasks were automated to speed the pace of innovation, reducing project times from days to hours.

To achieve these outcomes, MongoDB worked with Lombard Odier to create customizable generative AI tooling, including scripts and prompts tailored for the bank's unique tech stack, which accelerated its modernization by automating integration testing and code generation for seamless deployment.

Another way MongoDB helps organizations succeed with AI is by offering access to both technology partners and professional services expertise. For example, MongoDB has integrations with companies across the AI landscape—including leading tech companies (AWS, Google Cloud, Microsoft), system integrators (Capgemini), and innovators like Anthropic, LangChain, and Together AI.

[5] The importance of trustworthiness in AI-generated outputs can't be overstated. As McKinsey notes in a report on unlocking AI's potential, "data security, hallucinations, biased outputs, and misuse…are challenges that cannot be ignored."

An example of partners-plus-expertise success is CentralReach. CentralReach provides an AI-powered electronic medical record platform that's designed to improve outcomes for children and adults diagnosed with autism and related intellectual and developmental disabilities.

With MongoDB Atlas, CentralReach aggregated its diverse data types, so the company could build rich AI-assisted solutions on top of its database. Meanwhile, MongoDB partners like LlamaIndex helped CentralReach design and optimize multiple layers of its buildout. All of which will enable CentralReach to support initiatives like value-based outcome measurement, clinical supervision, and care delivery efficacy.

"As a mission-driven organization, CentralReach is always looking to innovate on behalf of the clinical professionals—and the more than 350,000 autism and IDD learners—that we serve globally," said Chris Sullens, CEO of CentralReach. "So being able to lean on MongoDBs database technology and draw on the collective expertise of MongoDB's partner network—in addition to MongoDB's tech expertise and services—to help us improve outcomes for our customers and their clients worldwide has been invaluable."

Adaptation (or Else)

In the AI era, what organizations need to do is abundantly clear: modernize and adapt, or risk being left behind.

Just look at smartphones, which have had an outsized impact on business and communication. For example, in its Q4 2007 report (which came out a few months after the first iPhone's released), Apple reported earnings of $6.22 billion, of which iPhone sales comprised less than 2%;[6] in Q1 2025, the company reported earnings of $124.3 billion, of which 56% was iPhone sales.[7] The mobile application market is now estimated to be in the hundreds of billions of dollars, and there are more smartphones than there are people in the world.[8] The rise of smartphones has also led to a huge increase in the number of people globally who use the internet.[9]

[6] https://www.apple.com/newsroom/2007/10/22Apple-Reports-Fourth-Quarter-Results/
[7] https://www.apple.com/newsroom/pdfs/fy2025-q1/FY25_Q1_Consolidated_Financial_Statements.pdf
[8] https://www.weforum.org/stories/2023/04/charted-there-are-more-phones-than-people-in-the-world
[9] https://ourworldindata.org/grapher/number-of-internet-users

However, saying "you need to adapt!" is much easier said than done. TDWI's research, therefore, is so important—it offers companies a road map for the future, and helps them answer the most pressing questions facing organizations as they confront the rise of AI.

How are your investments currently being spent, on maintaining dated infrastructure or creating customer value? Is your software agile enough to support AI development and AI-driven features without extensive retooling? And do you want to simply keep up with the changing world, or do you want to shape it?

To learn more about how MongoDB can help you create transformative, AI-powered experiences, check out MongoDB for Artificial Intelligence. And for more on the critical role that data plays in powering modern applications, check out MongoDB's digital magazine, Database Digest.

Transforming Data With Intelligence™

TDWI Research provides research and advice for data professionals worldwide. TDWI Research focuses exclusively on data management and analytics issues and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of the business and technical challenges surrounding the deployment and use of data management and analytics solutions. TDWI Research offers in-depth research reports, commentary, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

A Division of 1105 Media
6300 Canoga Avenue, Suite 1150
Woodland Hills, CA 91367

info@tdwi.org

tdwi.org