TDWI INSIGHT ACCELERATOR

Using Generative AI to Improve Operational Efficiency and Data-Driven Decision-Making

By James Kobielus

Sponsored by

snowflake®

tdwi

HOW GENERATIVE AI WILL TRANSFORM THE FUTURE OF WORK

Generative artificial intelligence (AI) is poised to transform practically every type of work. The potential of generative AI lies in its ability to automate development of every type of knowledge-based output. In the process, generative AI can boost the operational efficiency of many business processes that traditionally have relied on manual human efforts, as well as augmenting the productivity of humans in many knowledge-intensive functions.

Already, generative AI is being used to create source code, news articles, marketing materials, and document summaries. It can automate prompt-driven generation of images, video clips, musical compositions, and other non-textual outputs. Generative AI shows potential for fabricating a wide range of physical artifacts by driving devices in robotics, 3-D printing, manufacturing, and other domains, leading to greater operational productivity and efficiencies.

Generative AI applications leverage machine learning models to automate the generation of content in keeping with statistical patterns that are learned algorithmically from data in the relevant application domains. They do this by leveraging *foundation models*, which are large language models (LLMs) and other neural network algorithms that have been pretrained with semisupervised learning and incorporate sophisticated techniques such as transformer models and diffusion models.

As adoption of generative AI increases, interest in the technology continues to grow. In a recent TDWI survey, more than 30% of respondents plan to use generative AI in the coming year, in addition to the 20% who use it currently.

Early adopters of generative AI services are accessing them through SaaS or open source offerings. However, most companies TDWI speaks with prefer to rely on their own cloud data environments to run generative AI. This is due to privacy, security, compliance, and other considerations that require the controls associated with on-premises deployments. They may utilize a pre-trained model and bring it to their own environment to train it further.

This TDWI Insight Accelerator focuses on current enterprise practices and key steps your organization can take to unlock the potential of generative AI.

IDENTIFY KEY USE CASES IN OPERATIONAL EFFICIENCY

Generative AI is being incorporated into business analytics applications to drive creation of reports, dashboards, visualizations, and other data-driven outputs.

Indeed, business analytics is a primary place to look for high-value use cases of generative AI. In a recent TDWI survey, nearly 85% of respondents—of varying maturity levels in data and analytics practices—stated that their organizations have adopted reports, dashboards, and data visualizations as key business tools. A large proportion of those are also adopting predictive analytics, machine learning, and AI for operational analytics and decision support (see Figure 1).

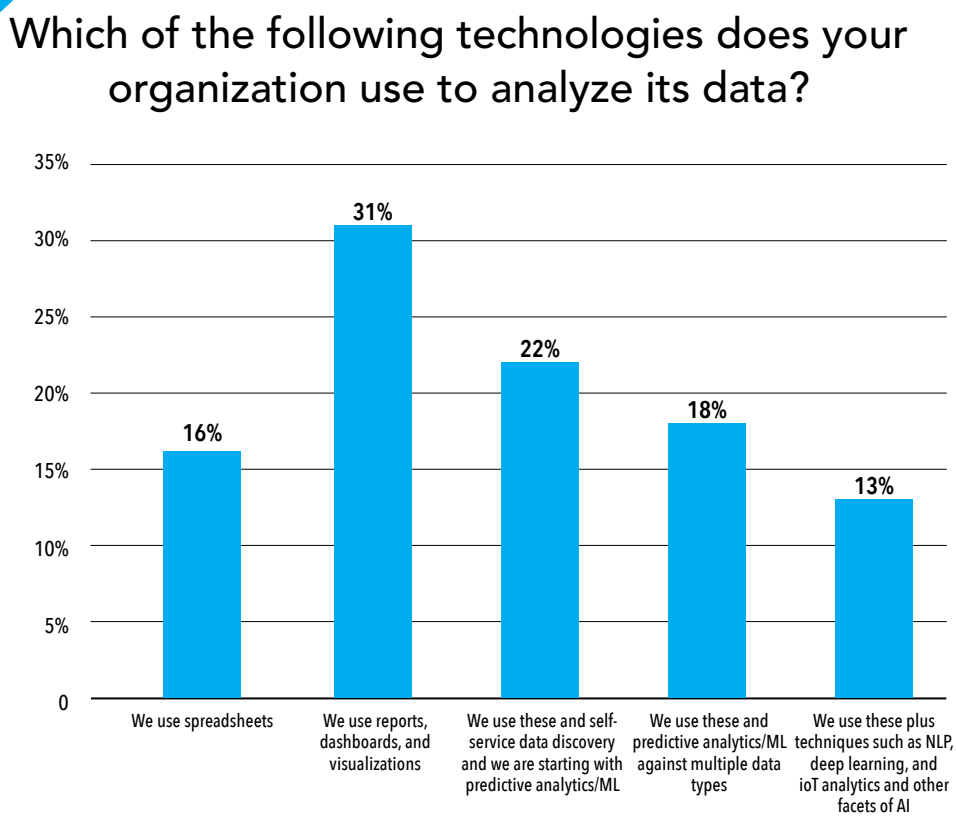

Figure 1. *Based on 119 respondents.*

Practically any search-driven analytics application can benefit from generative AI. For example, generative AI can improve natural language query so a category manager in a retail or consumer goods company can get fast answers about consumer sentiment to guide planning decisions. Applications can leverage LLMs to automate the creation of reports, dashboards, and other decision support outputs either on a scheduled basis or ad hoc in response to natural language queries. Enterprises can accelerate development of these generative AI–based analytics applications if they converge all their data into a core cloud-based environment that is their principal platform for operationalizing AI and ML.

Enterprises might implement generative AI in the form of NLP-driven digital assistants that deliver real-time decision support for various roles, tasks, and functions. Powering these digital assistants would be LLMs trained on the different enterprise data sets sourced from the corresponding business domains.

For example, an AI-powered customer service assistant could provide summaries of cross-channel customer interactions both for human agents as well as for the customers themselves via website chat functions. An IT service management AI assistant might provide help desk personnel with on-demand, up-to-the-moment summaries of problem status and incident histories. Still other

generative agents might roll up unified summaries that combine customer, product, financial, logistics, and other information for a unified view to manage workflows that cut across traditional business domains.

DEPLOY THE ENABLING DATA INFRASTRUCTURE

Unlocking the potential of generative AI requires enterprises to deploy a data infrastructure—either internal or through a SaaS or hosted offering—suited to its intended use cases.

For operational and decision support applications, enterprises will need to train and fine-tune generative foundation models with their own data. Enterprises will need to implement scalable data infrastructures for managing the unlabeled training data, algorithm libraries, and models associated with generative AI projects.

Organizations should focus on leveraging their data warehouses and other governed platforms for this purpose. In a recent TDWI survey, more than 80% of respondents stated that their organizations have adopted at least one of these governed data platforms, with a substantial portion of those respondents stating that they are using them as building blocks in a well-architected ecosystem for supporting more sophisticated data applications (see Figure 2).

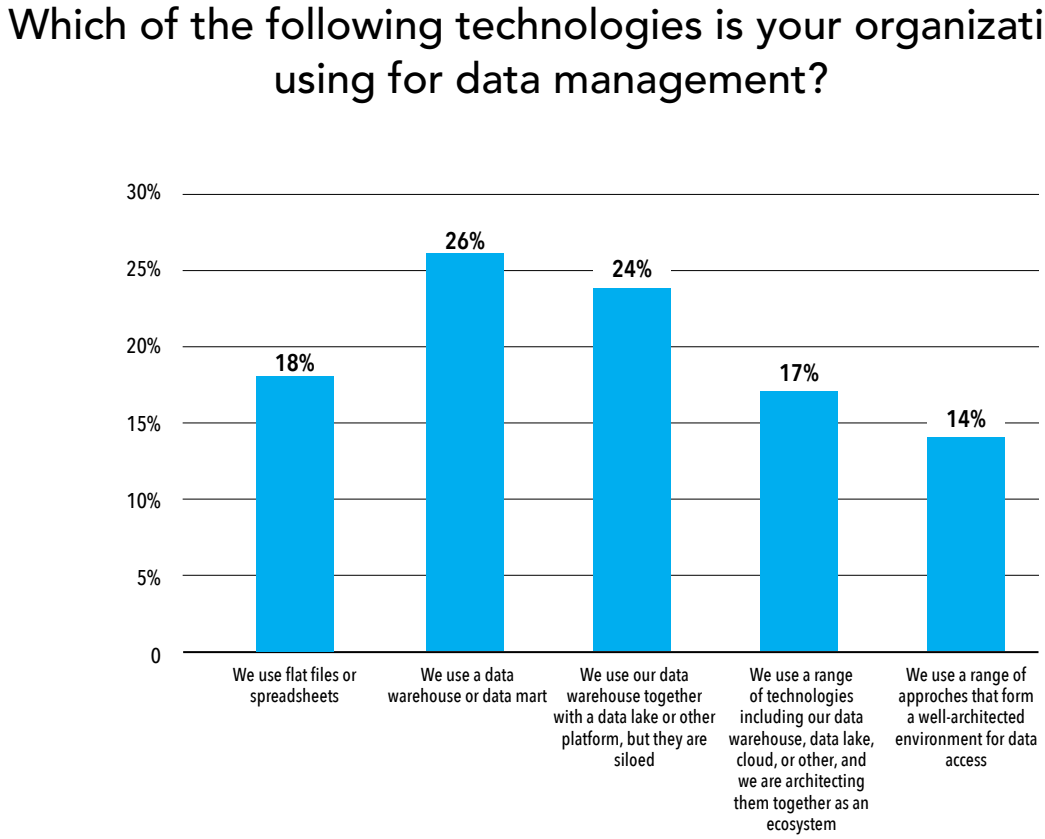

Figure 2. *Based on 119 respondents.*

SELECT AND TRAIN APPLICABLE FOUNDATION MODELS

The power of generative AI in business analytics and decision support comes from its ability to leverage foundation models that detect deep, nuanced statistical patterns in huge amounts of unlabeled training data.

Enterprises that implement generative AI must select the most suitable of many available foundation models for their specific requirements. Generative AI solution providers are also starting to provide commercial products that are designed and engineered for highly domain-specific, vertical-industry applications. Your enterprise will need to decide which foundation models best fit your requirements and how to bring those models into your own NLP, data science, and application development environments.

Organizations can only realize the benefits of foundation models for generative AI if they invest in platforms, tools, and staff to perform training of those models against data in the domain of the application, such as customer or logistics data. Foundation models are updated throughout the training process as they learn from data. Through training with data in the associated application domain, foundation models can continuously learn and iteratively generate newer, more suitable and optimal output. In some cases, they can be remarkably accurate in performing their tasks even when trained on little or no domain-tailored data. They can analyze a wide range of data—ranging from structured relational tables to unstructured text stores—to look for trends, categories, anomalies, and other patterns.

Fortunately, there are numerous pre-trained foundation models that enterprises can choose from when implementing generative AI. LLMs are foundation models that can predict which word statistically comes next within the response to a user-entered text prompt or instruction. Many are LLMs for text-to-text NLP use cases, and a growing range of foundation models are optimized for text-to-image and other multimodal applications. Pre-trained foundation models might be open source; others might be proprietary black-box models available only through APIs from hyperscale providers.

PROGRAM AND FINE-TUNE APPLICATIONS TO DELIVER THE BEST RESULTS

Generative AI requires tools for use by developers, data scientists, and domain specialists to write code, fine-tune foundation models, and prototype and refine working applications. Through no-code techniques, it enables business analysts to rapidly develop new analytics applications to support such use cases as optimizing logistics planning, maximizing marketing ROI, and evaluating consumer sentiment.

One of the core approaches in this regard is prompt engineering, which is an iterative process of submitting prompts to an LLM or other foundation model, reviewing the outputs, and then revising the prompts to bring the outputs in line with desired results. Through the back-and-forth sequence of prompts and outputs, generative AI applications may appear to be having a "conversation" with the user, which the user can guide by modifying, expanding, and refining the prompts until they receive the desired result.

Foundation models can be reused for new tasks beyond those for which they were initially trained. They can continuously learn from new data and iteratively generate newer, more suitable and optimal output. Rapid modification of

generative AI applications is possible through fine-tuning an LLM with new domain-specific training data, tweaking their machine learning weights to suit them for a new task, and supervising their outputs through reinforcement learning with human feedback.

For instance, if a marketing campaign isn't working, theoretically the foundation model it uses can be iteratively fine-tuned to address a new application requirement based on incoming results—such as generating a slightly tweaked marketing message suited to a different audience. Alternately, if a model has been pre-trained to translate call center scripts from English to French, it may—through training on a large set of documents in other, less closely related natural languages—be quickly adapted to also automate translation of text into those languages as well.

PROVIDE ONGOING GOVERNANCE OVER DATA, MODELS, CODE, AND OUTPUT

Organizations will need to implement data governance and curation to ensure that generative AI applications generate output that is trustworthy, relevant, compliant, bias-free, transparent, non-toxic, and otherwise suitable. For instance, if a generative AI application is providing targeted campaign messages in different themes and tones, these will have to be monitored for accuracy as well as bias.

Generative AI increases the need for humans—especially data and analytics professionals—to be in the loop in several areas integral to the governance of this technology. The technology requires strong supervision and curation over the data, models, and code used to implement these applications. Much of this will require a human-in-the-loop approach so that responsible parties (data stewards, privacy advocates, ethics supervisors, etc.) can do their jobs effectively. If generative AI is producing synthetic data, enterprises need processes in place to ensure that the data is suitable for its intended applications and doesn't leak private, sensitive, or otherwise controlled data.

Organizations will need to put standards and rules in place quickly to address these evolving governance requirements. Additionally, organizations will need to revisit how generative AI is impacted by regulations such as GDPR in terms of privacy and transparency.

RECOMMENDATIONS

To accelerate deployment of high-value operational efficiency and decision support applications, TDWI believes that enterprises should undertake the following generative AI journey.

Identify compelling use cases for generative AI.

Moving forward with generative AI will first involve determining what business imperatives are best addressed with this technology. Once you identify priority use cases—such as providing a generative AI–powered chatbot for making product recommendations—you will need to pinpoint the data necessary to build, train, and operationalize the application. You must determine if the data exists, if you have rights to use it, if it's trustworthy and relevant to the intended application, and if it's in the proper format and schema. If you're building operational and decision support applications of generative AI, focus on using existing data from your enterprise data warehouse, data lake, and other governed repositories.

Develop a generative AI proof of concept.

After prioritizing initial generative AI use cases, develop a prototype as a proof of concept. This should involve training the appropriate foundation model with the relevant data, supervising the model output, and assuring it is accurate, appropriate, and relevant. It also involves testing the prototype with users and determining the extent to which they find it useful. This is a highly iterative process that should be aligned with enterprise DevOps, agile, and other operationalization practices. If you're building operational and decision support applications of generative AI, you should focus on using foundation models that provide text-to-text capabilities for generating optimal queries, reports, and summaries of operational data. You should also use foundation models for text-to-image generation of operational dashboards, charts, and data visualizations.

Operationalize the generative AI application.

Once the generative AI prototype has been validated in the field, operationalize it in accordance with established practices. This will involve provisioning necessary storage, processing, and other IT resources; defining governance, security, and other policies and practices; and using observability tools to monitor the application in production. If productionizing operational and decision support applications of generative AI, a key focus should be on establishing the necessary data and model governance workflows consistent with existing enterprise practices for business analytics and data science. Data and analytics professionals will need to be educated about how to train a generative foundation model, how to build high-quality applications that leverage the model, and how to monitor, administer, and secure it in production.

Educate and train users on the new generative AI applications, making them aware of the productivity benefits to their various jobs and sensitive to the risks of abuse or misuse of the technology.

ABOUT OUR SPONSOR

snowflake®

Snowflake enables every organization to mobilize their data with Snowflake's Data Cloud. Customers use the Data Cloud to unite siloed data, discover and securely share data, power data applications, and execute diverse AI/ML and analytic workloads. Wherever data or users live, Snowflake delivers a single data experience that spans multiple clouds and geographies. Thousands of customers across many industries, including 639 of the 2023 *Forbes* Global 2000 (G2K) as of July 31, 2023, use Snowflake Data Cloud to power their businesses.

Learn more at snowflake.com.

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.

ABOUT THE AUTHOR

James Kobielus is senior director of research for data management at TDWI. He is a veteran industry analyst, consultant, author, speaker, and blogger in analytics and data management. He focuses on advanced analytics, artificial intelligence, and cloud computing. Kobielus has held positions at Futurum Research, SiliconANGLE Wikibon, Forrester Research, Current Analysis, and the Burton Group and also served as senior program director, product marketing for big data analytics, for IBM, where he was both a subject matter expert and a strategist on thought leadership and content marketing programs targeted at the data science community. You can reach him by email (jkobielus@tdwi.org) on X/Twitter (@jameskobielus) and on LinkedIn (https://www.linkedin.com/in/jameskobielus/).