

IT Operations

Put AI to work



IBM

Contents

Today's IT teams work in a hybrid IT landscape: cloud, on premises systems and edge devices. Implementing, managing and delivering IT operations across these complex environments can be challenging, but generative AI (genAI) can help.

Let's explore how.

IT in the age of AI

[03 →](#)

Put AI to work for IT

[06 →](#)

Next steps

[20 →](#)

Supplement:
GenAI for FinOps

[27 →](#)

IBM can help

[31 →](#)

01

IT in the age of AI

Business success depends
on efficient *and* effective
IT operations.

In our highly digitized world, both businesses and customers expect applications and systems to perform optimally. And downtime is expensive—costing an organization close to USD 1 million a year on average according to an IDC study.¹ That issue makes the health and performance of applications and the systems they run on a priority, not just for IT teams and CIOs, but also for the entire C-suite.

However, today's IT operations are increasingly complicated. They span hybrid IT environments, with applications and workloads running across multiple clouds, on premises and at the edge. IT teams must implement, manage and deliver

services across these diverse platforms—using disparate tools and systems—while addressing performance and integration issues, compliance and security risks, and cost management challenges. No surprise, it's a struggle for most IT teams.

Yet, as IT environments evolve, IT operations should evolve too.

GenAI offers IT—its teams, tools and processes—tremendous opportunities for transformation. IT, enhanced by genAI, has the power to be a key driver for greater efficiency, productivity and performance. It can also help deliver needed innovation and responsible stewardship and bring about seismic cultural shifts as well.

There are challenges, of course. Delivering ROI, managing employee expectations and creating trustworthy systems for ethics and governance won't be easy. Yet the risks of maintaining a stale status quo—or worse, falling behind the competition—will be harder to handle.

GenAI leaders are already realizing real value from their efforts: 72% greater annual net profits, and 17% more annual revenue growth than their peers.²

87%

say the technology helps them execute more high-impact automation initiatives.

74%

are investing in AI-powered automation to decrease IT and network complexity.

66%

state that predictive AI capabilities will detect problems earlier.

58%

are investing in AI-powered automation to deliver new and improved IT platforms, applications and networks.



2

Put AI to work for IT

AI in IT isn't new as enterprises try to handle the growing complexity of modern IT environments and the data created by them.

Using AI for IT operations (AIOps), IT teams work to:

- Understand the applications and systems in their current state to address potential incidents.
- Accelerate resolution to help minimize any downtime that impacts the business in the event of an incident.
- Increase efficiencies by reducing manual efforts, mundane tasks and toil.
- Control costs better with trustworthy, automated actions at speed and scale.
- Optimize the overall use and spend of tech, including infrastructure, to drive business value.

In the past, enterprises have approached AI as an add-on. Now, **AI is becoming the centerpiece of business transformation**—75% of business leaders surveyed believe competitive advantage will depend on who has the most advanced genAI.³ But fundamentally transforming IT operations requires a mix of vision and technology with AI operating at the strategic core of the business.

So, does this mean you need to replace your traditional AI solutions with the latest genAI for IT operations? IBM AI leaders say no.



Senior executives are trying to save money through optimization. They don't want to cut headcount; they want more productivity from what they already have. That's what digital transformation is. If I want to be more efficient, I just work harder. But if I have technology and tools to offload some of it and machine learning to make recommendations ... that's being more productive.

William Lobig

Vice President, Product Management,
IBM Automation Software

Traditional AI

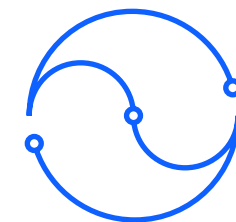
Traditional AI models that use conventional machine learning (ML) and rules-based models have different capabilities and serve different functions: predictive analysis, security and compliance, automation, and more.

GenAI models, on the other hand, use [foundation models](#) to autonomously generate content based on the data the models were trained on. They're complementary technologies that you can use together.

Applies predefined rules and algorithms to specific sets of data to help solve problems, make predictions and automate tasks.

Potential IT tasks using traditional AI

- Predictive maintenance
- Problem detection
- Incident response
- Service desk automation
- Process automation
- Intelligent workflows
- Capacity planning



Generative AI

Uses both small language models (SLMs) and large language models (LLMs) to generate new content based on patterns learned from the data they were trained on.

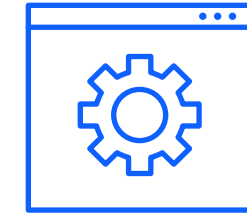
Potential IT tasks using genAI

- Issue summarization
- Action recommendations
- In-depth conversation
- Code generation
- Insight extraction
- Classification
- Testing and debugging



With genAI and traditional AI, IT teams can more easily move from a *break-fix* model to a more proactive one that enables predictive, preventive and recommended actions. AI offers opportunities at a scale that simply can't be matched by humans—no matter how skilled or experienced.

Strategic use cases that IT teams should consider to help boost IT operations and development include:

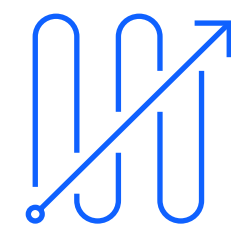


Resolve application issues before users are impacted.

Use real-time root cause analysis capabilities—powered by AI and intelligent automation—to swiftly identify the underlying causes of incidents. Then take immediate action to reduce both mean time to detect (MTTD) and mean time to resolve (MTTR).

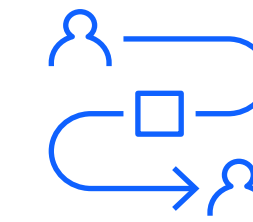
Streamline IT planning, budgeting and forecasting.

Shift from data preparation to data-driven analysis, and proactively manage IT and labor spend against budget. Use AI to align resources with strategic priorities and optimize investments in technology.



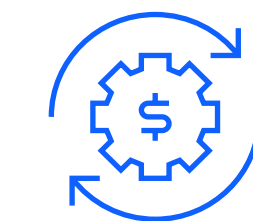
Guide users through automated discovery.

Equip users to generate insightful analysis and recommendations through a natural language interface. Users can ask questions such as *what is the root cause?* or *what is the best course of action to address it?* and evaluate the impact of decisions.



Optimize costs.

Use genAI to accelerate the execution and utility of your hybrid cloud model to better control costs. Use the public cloud for temporary or bursty workloads, paying for resources— including those for AI-based projects— only when needed. Manage long-term or sensitive workloads on premises or in a private cloud to maintain cost efficiency.

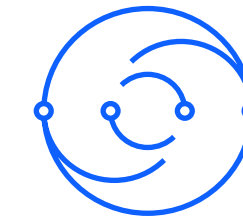


Augment IT teams with AI assistance.

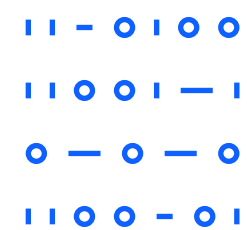
Use assistants to help teams troubleshoot and resolve issues, with quick access to relevant information, and automate repetitive tasks—such as data entry, scheduling and monitoring. Also use the assistants to identify patterns and anomalies to help IT teams optimize systems and processes.

**Redesign data integration.**

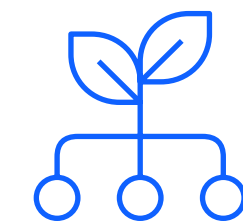
Use generative AI to reduce the time needed to connect applications and systems and unlock critical data.

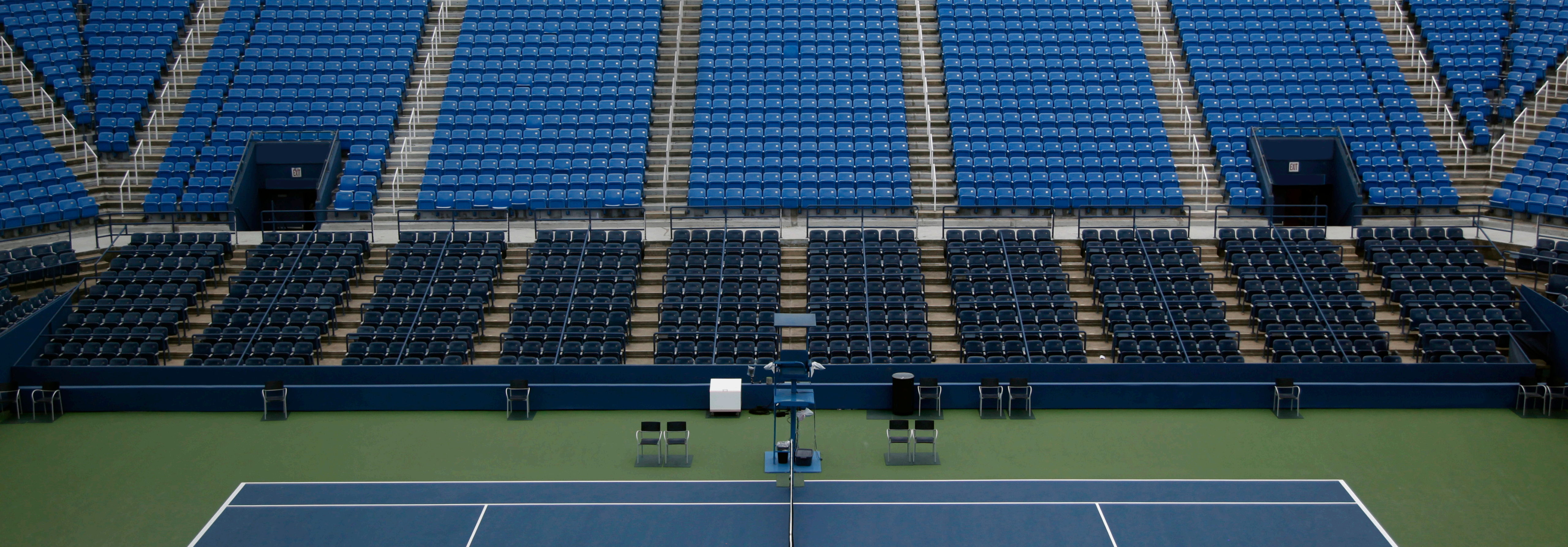
**Enhance code generation.**

Increase developer productivity with AI-generated code recommendations based on natural language requests or existing source code. Extend the process to playbook creation, infrastructure provisioning and anything else that can be deployed as code.

**Create more sustainable IT.**

Analyze resource usage data and use automated actions to identify, report on and reduce carbon emissions. Use genAI to identify patterns of potential waste.





GenAI in action

Many organizations around the world are already using genAI to optimize their IT operations and management. Here are two IBM clients that found success when they put AI to work for their IT operations.

To help the [US Open](#) stay on the cutting edge of customer experience, IBM Consulting® worked closely with the USTA to develop genAI models that transform tennis data into insights and original content on the US Open app and website. The USTA used the [IBM watsonx™ platform](#) to build the AI models supporting key app features such as Match Insights and the new AI Commentary for US Open highlight reels, as well as [IBM Instana® Observability](#) technology to monitor application performance and surface issues.

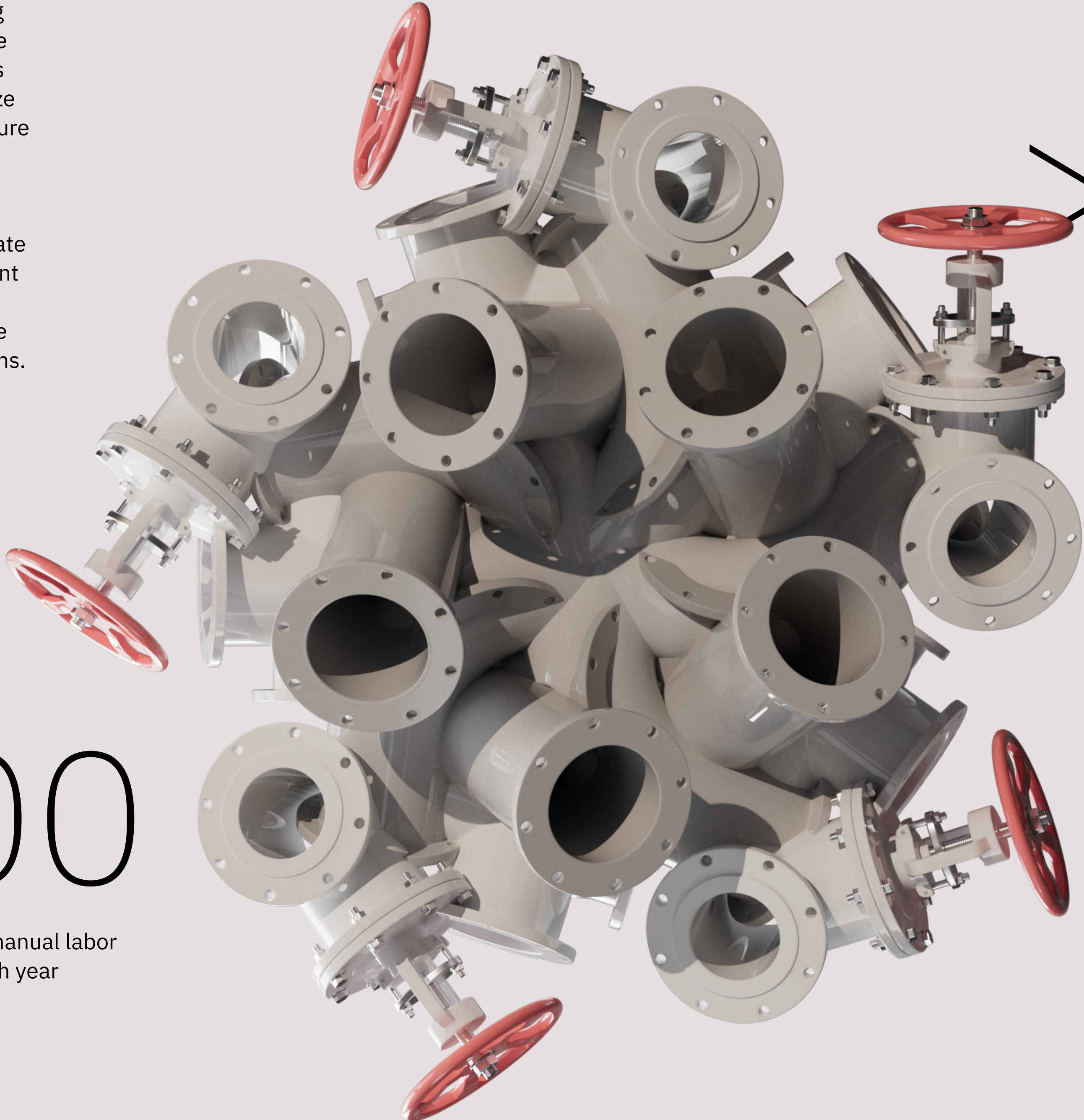
15M

world-class digital experiences for more than 15 million fans around the globe

7M

data points captured and analyzed by IBM throughout the tournament

In the dry climes of Western Australia, finding and maintaining clean, drinkable water can be challenging. Fortunately, [Water Corporation](#) is up to the task. To help the company modernize and migrate its mission-critical SAP architecture to the Amazon Web Services (AWS) cloud, IBM Consulting® put forward an agile and automation-first strategy. Water Corporation used Red Hat® Ansible® Lightspeed to automate the configuration, provisioning and deployment of its SAP systems and SQL databases and [IBM® watsonx Code Assistant™](#) to create code recommendations for the automation functions.



~ 1,500

hours of manual labor saved each year

> 40%

reduction in cloud-related operating costs

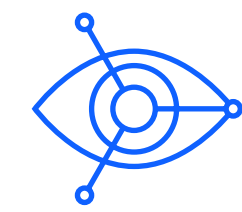
0

unplanned downtime over a one-year migration that included over 1,000 individual process steps

Applying genAI for IT operations

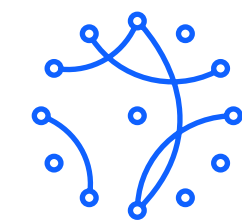
IT processes that are prime for genAI to augment, accelerate and advance can be generally categorized under **four phases: observe, engage, act and optimize.**

What happens in these phases?



Observe

Correlate and contextualize all real-time and historical data across hybrid IT environments to detect anomalies, reduce noise and focus on issues that matter.



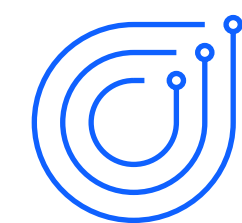
Engage

Alert and route incidents to appropriate teams. Highlight key information and recommendations at user touchpoints to quickly address issues and diagnose and identify root cause.



Act

Automate and orchestrate remediations to quickly resolve issues.



Optimize

Enhance IT performance and cost efficiency continuously with ongoing insights and improvements.

Which key processes align with these phases?

- Data collection and analysis
- Incident detection and scope analysis
- Report generation and visualizations

- Incident and problem management
- Automated ticket translation and triaging—clustering, classification and prioritization
- Knowledge management and assistance—summarizations, documentation and recommendations

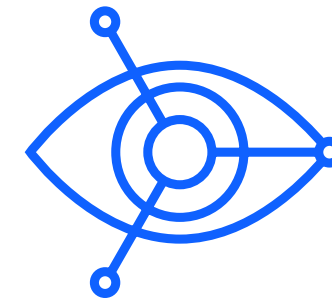
- Automation playbook creation
- Change management

- Cost and capacity tracking and analysis
- Tech investments-to-business outcomes alignment

Let's take a deeper dive into each of the phases and processes to understand where IT operations can benefit most from genAI. ↓

Observe

Nearly two-thirds of responding executives that believe automation is essential to fast-track genAI capabilities are already automating topology discovery (66%), incident mitigation (69%), and performance monitoring (63%).⁴



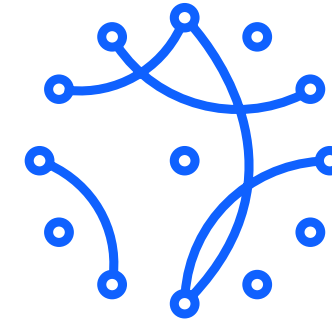
Improving end-to-end observability is essential to autonomous IT operations; so you can stay informed about what's happening across the entire hybrid IT estate. Integrating genAI into observability further simplifies access to the most critical insights, making it easier to identify and locate critical events as they happen across your applications, data and processes.

- Enhance accessibility of observability dashboards—expanding usability across roles with simplified query interfaces.
- Automate report generation with descriptive summaries or visualizations—helping provide deeper insights into system behavior and performance trends.
- Detect emerging incidents more effectively—highlighting IT operational issues with a higher degree of confidence.
- Improve understanding of incident scope—enabling conversational what if analysis with situational awareness of the IT environment.

[Explore IBM Observability solution →](#)

Engage

85% of responding executives agree genAI is enabling digital assistants to make predictions and generate solution ideas to complex problems.²

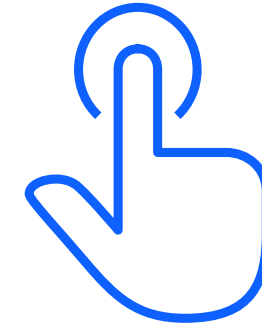


GenAI can dramatically streamline IT service management and measurably improve processes, such as ticket triaging, incident prioritization and issue summarization, to further save time when responding to—and even mitigating—urgent issues.

- Improve alerting with more rapid and active communication and more effective messaging.
- Identify and group incidents with similar root causes more efficiently and route to the appropriate teams with recommended actions.
- Enhance knowledge management and discoverability with auto-generated summaries for documenting incident highlights and actions.
- Augment teams with genAI-powered assistance, conversational access and auto-translation to streamline and optimize the incident management process.

Act

62% of IT executives say their organizations are using genAI for code generation—and that figure is expected to jump to 87% by 2026.⁴

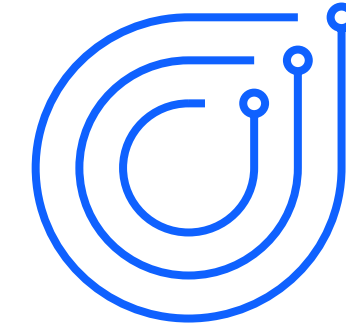


Another popular use for genAI is to help create code with natural language prompts. A genAI-led incident response automation could quickly generate code that creates changes to applications and infrastructure and remediate a performance issue with minimal human intervention.

- Generate automation playbooks to expedite fixes and incident resolution.
- Convert standard operating procedure into answerable code.
- Generate infrastructure code, including Terraform infrastructure code, with natural language prompts.

Optimize

57% of IT executives are already using genAI to predict their ROI on IT investments—and this figure is expected to grow to 75% by 2026.⁴



With greater adoption of hybrid cloud architectures, achieving cost optimization without compromising service quality has never been greater. With AI, systems are continuously learning to deliver better, faster insights and automations. The application of genAI can help amplify those efforts and maximize tech spend—including that of genAI—to fast-track business value in the following ways:

- Discover telemetry and instrumentation gaps—such as over or under provisioning—and identify steps to correct them with more granularity and easier access through natural language querying.
- Improve AI workload placement to increase cost efficiency.
- Better understand the value of genAI investments aligned to business objectives.

OS3

Next steps

As enterprises transition from the *hype* to the *how* of genAI, they realize the need to tread carefully, but they also feel the need to act fast.

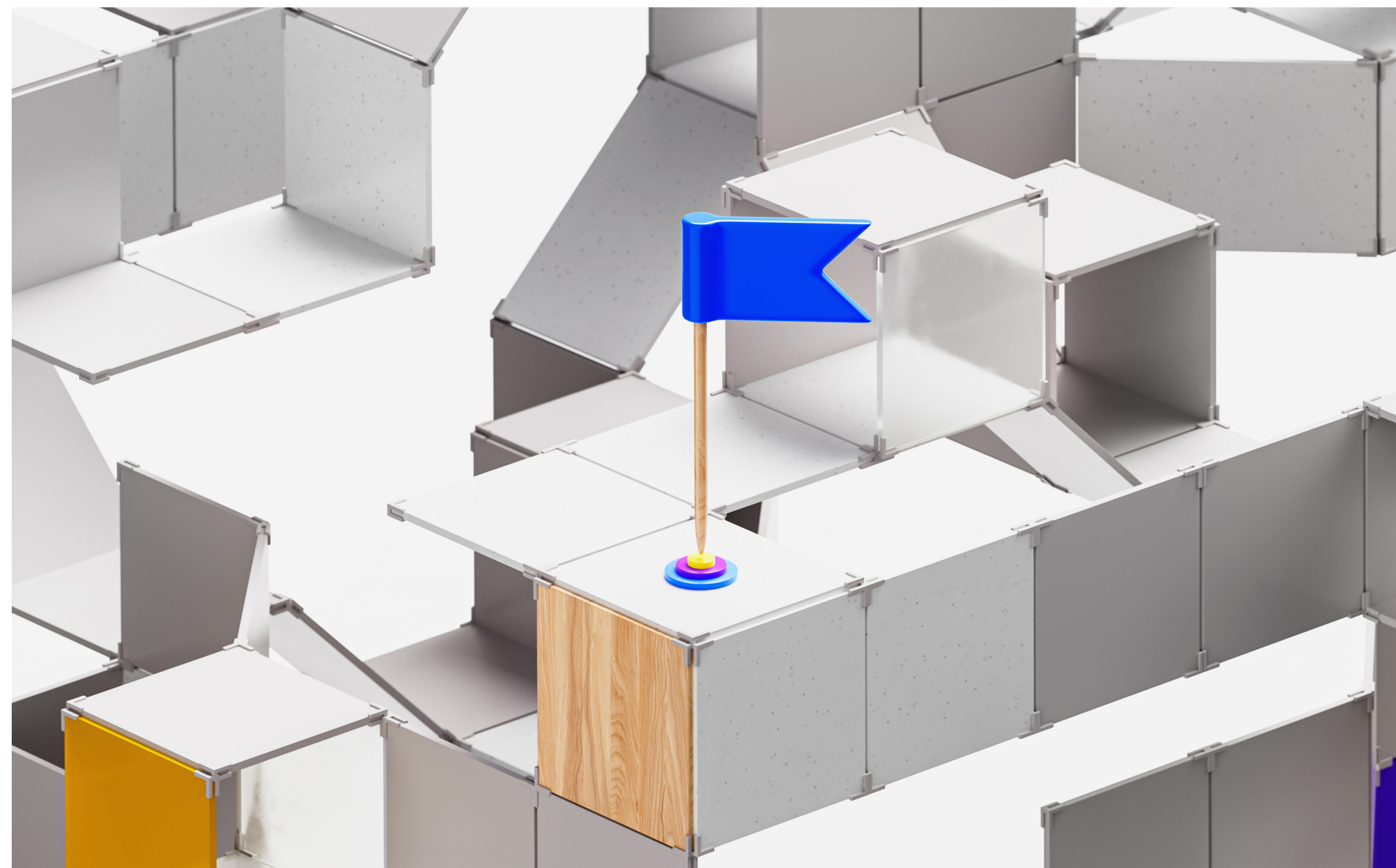
60% of organizations surveyed have not yet developed a consistent, enterprise-wide approach to genAI.⁶

Here's how you can get started [↘](#)

01 Create a small, achievable goal

Any good strategy starts with a clear understanding of what you're trying to accomplish. With genAI, the first step is to think small, but strategically.

Start with a documented set of processes or functions that your organization performs on a regular basis. Then choose one that touches large parts of your operations. That normally means it's a process that involves a lot of manual work and, hence, is ripe for automation and augmentation with AI. For example, the process to onboard a new customer may require interactions with 10 different people, departments and systems. Look at the effort required at each step and determine which of those 10 interactions will represent the highest-value opportunity for your organization if you speed them up or add in natural language capabilities to gain valuable insights. Think in terms of the resources or time that can be saved if you can readily get information and insights, such as recommendations on how to optimize the process.

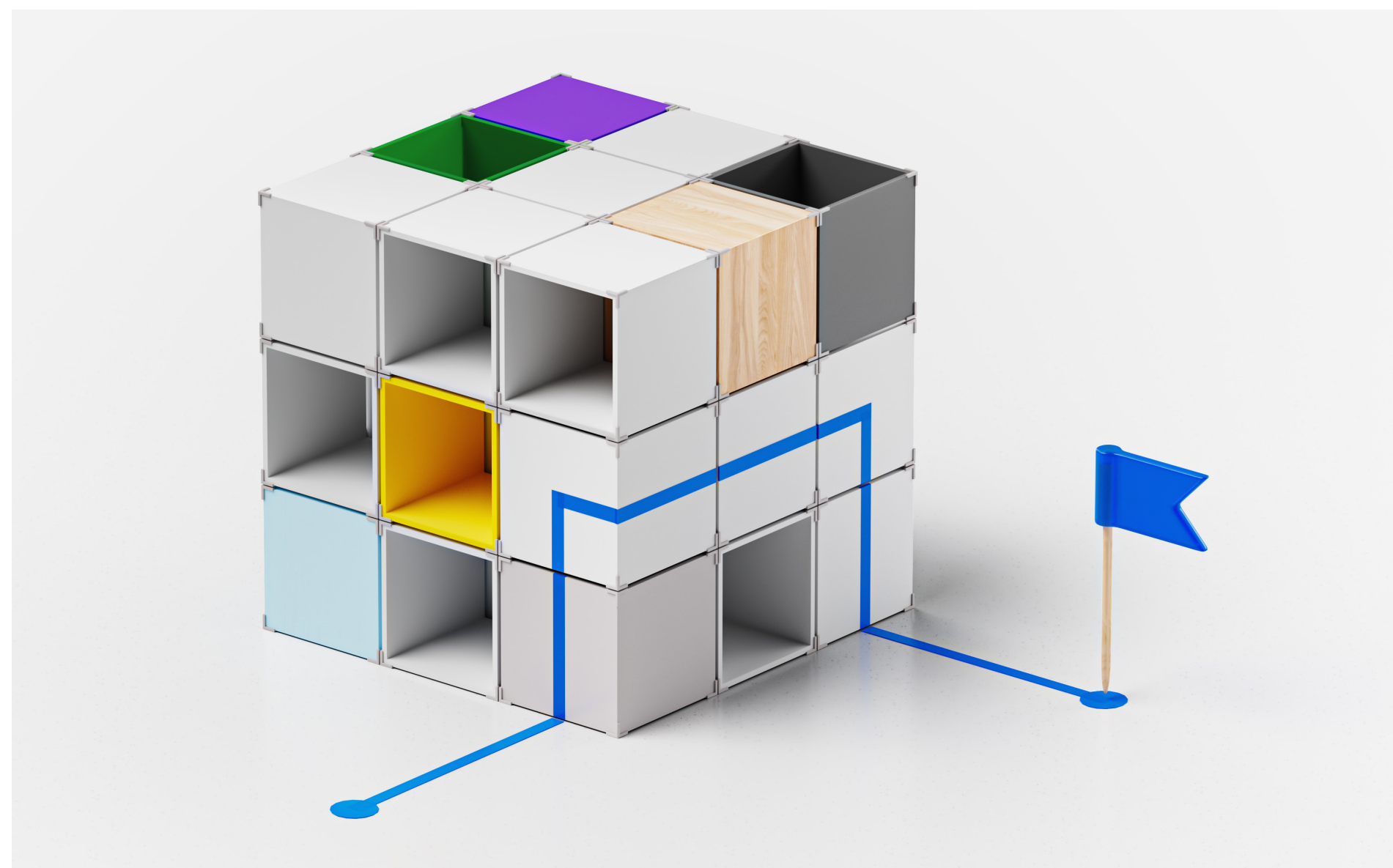


02 Determine metrics for success

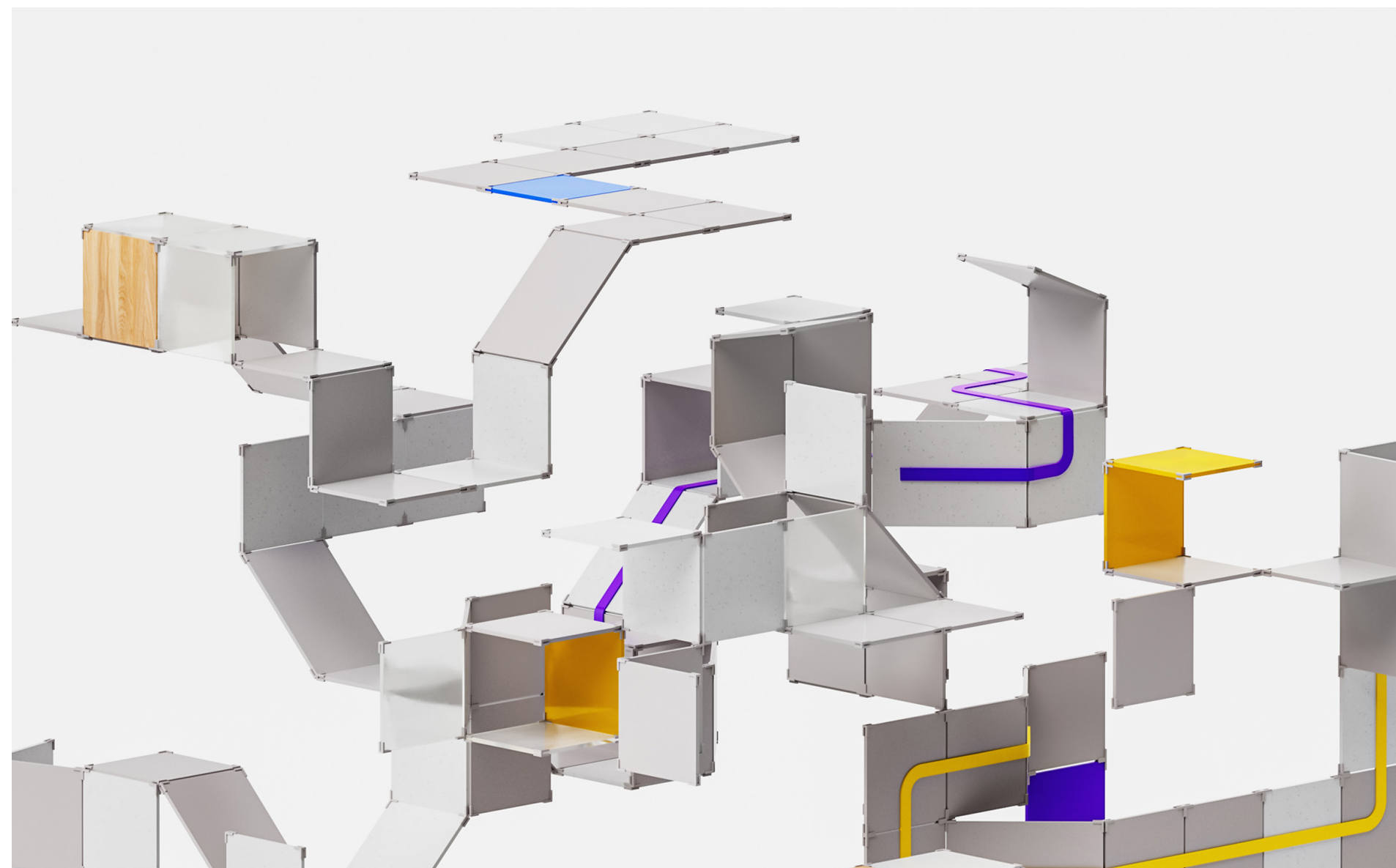
Once you've identified a starting point, you need to define how you'll measure success. When it comes to IT, key performance indicators (KPIs), such as mean time to repair and mean time to respond (MTTR) and application uptime, are common measurements. Using the onboarding example, you may set a goal of 99.99% availability for an online customer request form.

Beyond these standard metrics, other measurements are important but often harder to calculate. For example, how satisfied are customers with the new online form? As you infuse AI and automation into the mix, you may also consider pulling in metrics, such as net promoter scores (NPS) or online reviews. If you're trying to measure employee productivity or morale, consider using information from employee engagement surveys.

Change management is another area that's crucial but may be difficult to measure. For example, how much did you save by releasing a new feature ahead of schedule? How many new customers did you attract with those releases? Again, sources of data, such as brand health or NPS, will help you get a more accurate picture of the success of your IT automation investments.



03 Ask important questions about your data and infrastructure



An important consideration is your data itself—where it's stored, how to access it, how to keep it secure. And if quality data is key to effective AI, the infrastructure used to access and manage that data is even more critical. Especially the level of trust and transparency that's required across environments. Being data focused means bringing AI to the data, wherever it resides, quickly.

To do this process, you need to enable hybrid data access across multiple environments. Start with a strategy across the full stack to provide an optimized environment for AI deployments. This method will increase the efficiency of data access and management, to workload placement and management, including container platforms, developer platforms, ITOps and security.

And finally, work with what you have by making more intentional choices. Understand what outcomes you want to achieve and the IT architecture that can help you get there.

Having the right mix of on-premises and cloud infrastructure to support the requirements of AI workloads by design can ease these complexities significantly. It enables you to run and manage workloads wherever you want, no matter where the data resides.

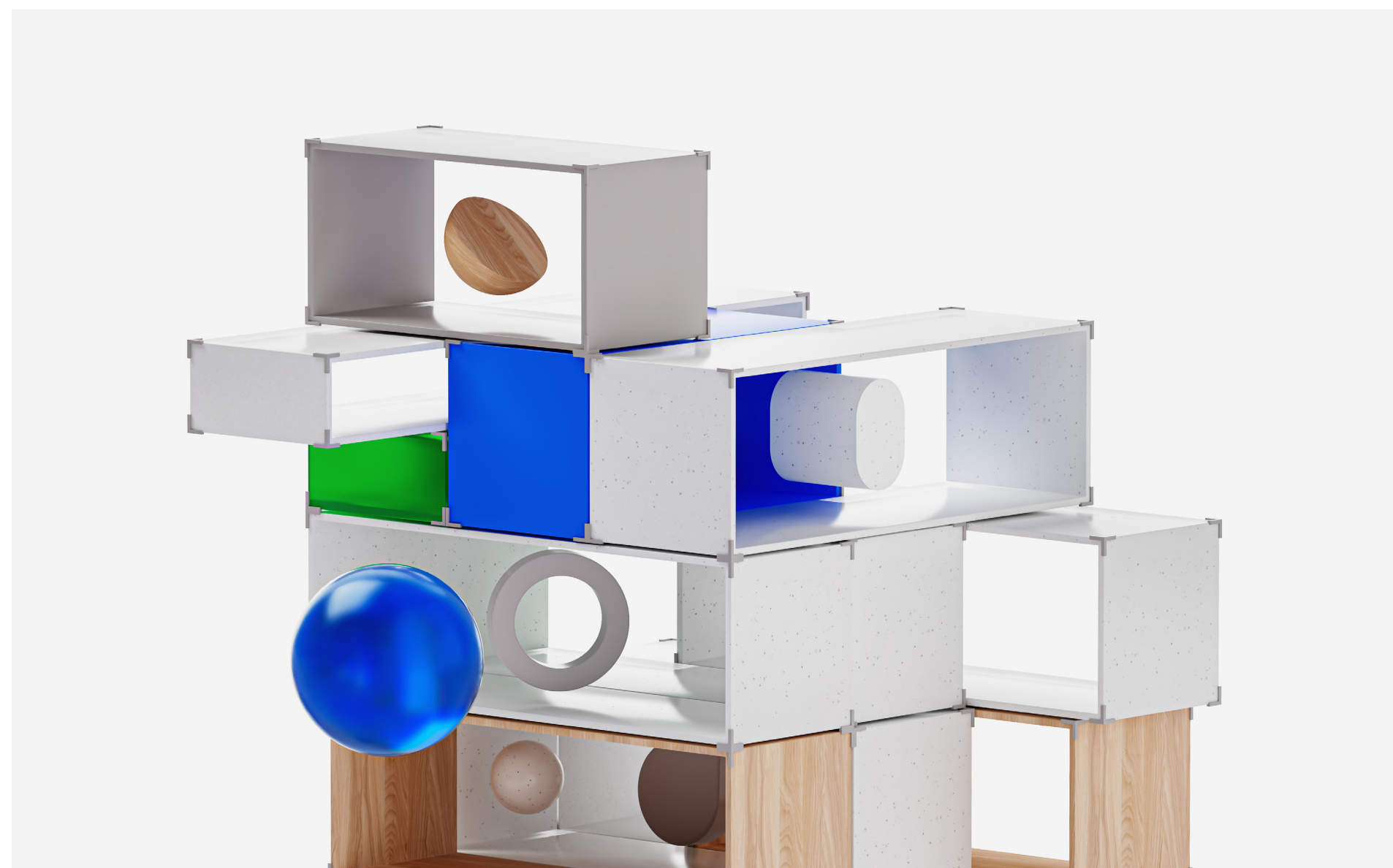
04 Evaluate the genAI model for fit

Other important factors to consider are the purpose of your application of genAI and the [foundation models](#) it's trained on. The model landscape is rapidly evolving, and you must balance several attributes, including model size, license type, potential risks and total cost. Do you need a massive model to execute your task? The answer, in most cases, is no.

General-purpose LLMs can be inefficient and lead to astronomical compute costs and energy consumption. And they most likely won't be trained to meet the unique requirements of your use case, business needs or domain.

Keep the following steps—and questions—in mind when choosing your genAI solution:

- Articulate your use case clearly.
What is it that you need genAI to do and is there a model that closely matches your unique requirements? Can you train a model on your enterprise data?
- Evaluate the model size, performance and risks.
Is the model right-sized for your specific application and use case? Is the way in which it was developed and sourced transparent? How is risk mitigated?
- Test options and choose the one that provides the most value.
Does it meet your cost and deployment needs? Does it help you reach your sustainability goals?

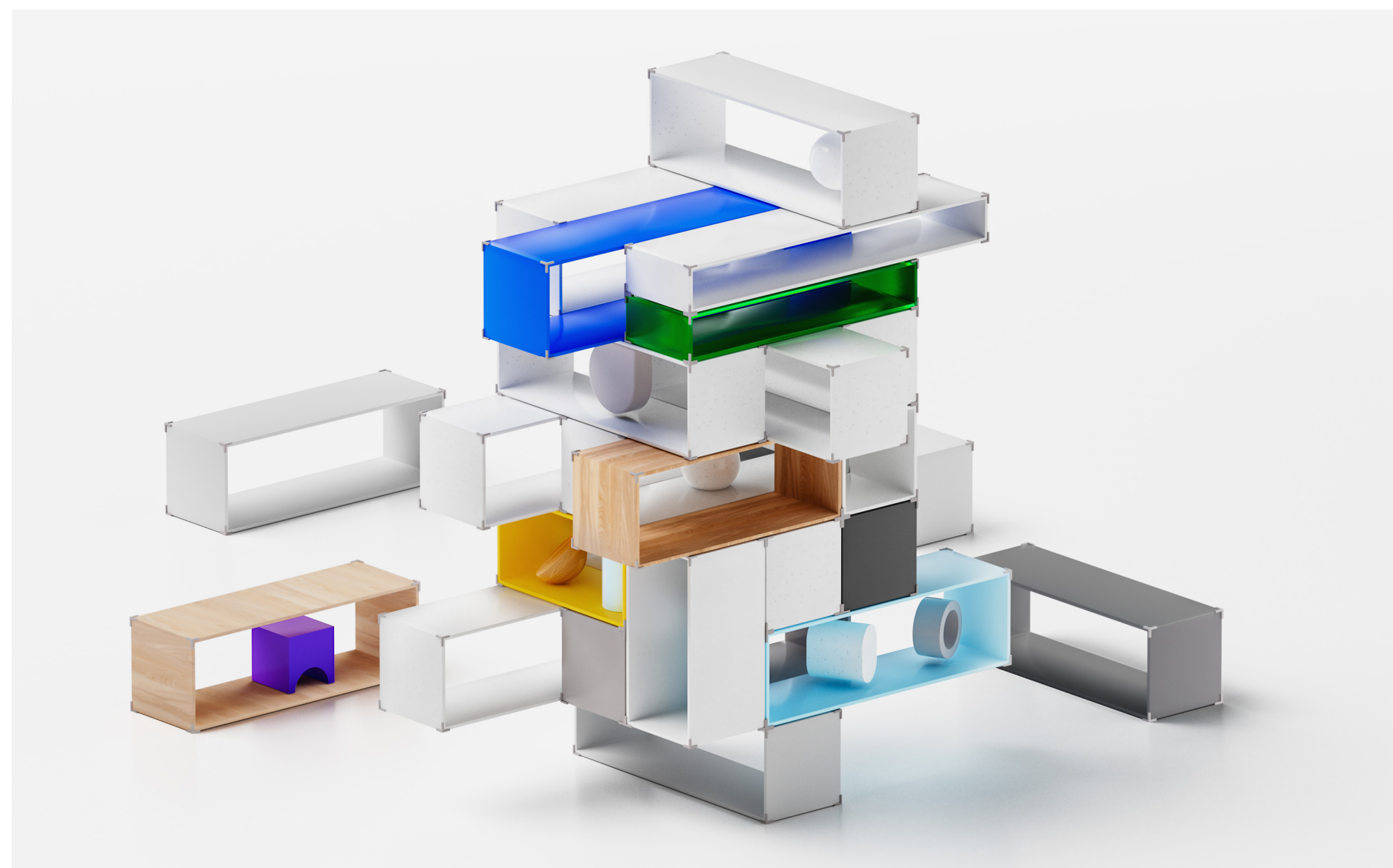


05 Understand the cost, then reinvent for innovation

As with most things related to IT, cost is always a factor—this factor is especially true as you consider adding AI and graphics processing unit (GPU) usage to the mix. However, only 15% of surveyed tech executives expect to fund any AI-related budget uptick with net-new spend.⁵ Instead, a third stated that money will come from non-AI IT spend.⁵ But reallocating money doesn't consider other costs—particularly labor and cloud spending, both of which will increase as demand for genAI solutions grows. For example, how many AI prompt engineers, ML engineers, AI data scientists, AI trainers and AI ethicists do you have on your staff now and what have you budgeted for the future?

That's why it's critical to focus on investments that truly pay off and deliver differentiation. Right now, surveyed CEOs are focused on cutting costs with 74% of genAI money going toward HR, finance, customer service, sales and marketing, and IT.⁵

Cost savings are crucial, but you can't economize your way to innovation. Instead, rethink your approach to bigger IT investments. Start by building an augmented FinOps practice. Once you have transparency and accountability for critical cloud costs, radically modernize your methods of evaluation. Instead of measuring short-term savings, consider the overall growth potential of your AI spend.



Supplement

Managing tech spend in the age of genAI with FinOps

Worldwide IT spending is expected to total **USD 5 trillion** in 2024, up almost 7% from 2023.⁷ Far too often, enterprises overspend on applications—around 24%—to make sure everything performs as expected. That’s money that could be better spent on innovation, infrastructure, security and better customer experiences.⁸

To maximize business value and realign tech spend, enterprises need visibility into costs and spending across all their hybrid IT investments. Enter FinOps, which brings together engineering, finance and technology teams to help you make data-driven spending decisions.



According to the FinOps Foundation:

“**FinOps** is an operational framework and cultural practice that maximizes the business value of cloud, enables timely data-driven decision-making and creates financial accountability through collaboration between engineering, finance and business teams.”

According to Gartner:

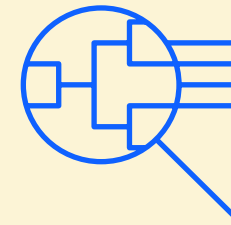
“**Augmented FinOps** automates FinOps processes through the application of artificial intelligence (AI) and machine learning (ML) practices—predominantly in the cloud—to enable environments that automatically optimize cost based on defined business objectives expressed in natural language.”

With FinOps, enterprises have the systems, best practices and team alignment to better understand cloud costs and make informed trade-offs.

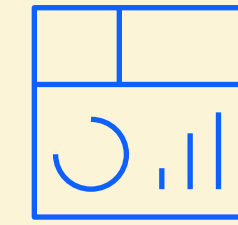
With it, apps that provide the most business value—as defined by your company—get the most support, while low-performing apps don't take up precious and costly IT resources.

When FinOps is augmented by genAI, organizations can get even more value with less effort, as the following use cases illustrate:

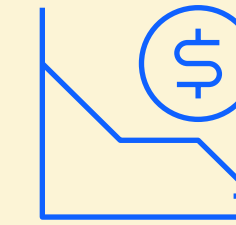
Predictive analytics: By analyzing financial data alongside market trends and business performance metrics, genAI can provide predictive insights into future financial outcomes. These insights can help organizations anticipate revenue streams, forecast expenses and make proactive decisions to mitigate risks and capitalize on opportunities.



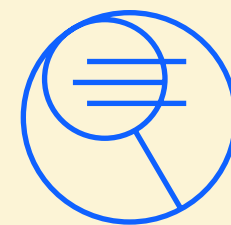
Automated reporting: GenAI can automate the generation of financial reports by extracting relevant data from disparate sources and synthesizing it into comprehensive dashboards or presentations. This method streamlines the reporting process, reduces manual errors and provides stakeholders with timely insights to support decision-making.



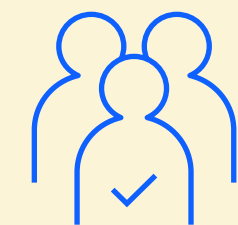
Cost optimization: GenAI can analyze historical financial data and current spending patterns to identify areas where cost savings can be achieved. It can recommend adjustments to resource allocation, such as optimizing cloud service usage or renegotiating vendor contracts, to minimize unnecessary expenditures.



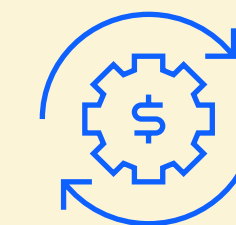
Compliance monitoring: GenAI can assist in managing compliance with financial regulations and internal policies by continuously monitoring transactions and financial activities for adherence to established guidelines. It can flag potential compliance breaches and provide recommendations for remedial actions to help maintain regulatory compliance.



Customer segmentation and pricing optimization: GenAI can analyze customer data to segment clients based on their profitability and behavior. It can then recommend personalized pricing strategies or incentives to maximize revenue while helping to ensure customer satisfaction and retention.



FinOps lifecycle: FinOps is designed so that practitioners work iteratively on the [FinOps Foundation Framework Capabilities](#) through three phases: Inform, Optimize and Operate.



Want to learn more about FinOps? Visit the [FinOps Foundation](#).

Inform: Visibility and allocation

Begin by empowering all stakeholders with the information and understanding they need to make informed, cost-effective decisions around cloud usage.

Start with the fundamentals:

- Insights into multicloud cost and usage from cloud service providers
- Cost allocation
- Business mappings
- Reporting and dashboards

Then tackle more advanced tasks:

- Cloud program total cost of ownership TCO and unit economics
- Unit costing
- Advanced allocation of shared costs
- Profitability and chargebacks

Optimize: Rates and usage

Next, identify opportunities for additional savings and improve cloud efficiency using the data and capabilities developed in the *Inform* phase.

Start with the fundamentals:

- Rightsizing recommendations and preferences
- Container cost optimization
- Reserved Instances (RI) and Savings Plans (SP) recommendations
- RI planner

Then tackle more advanced tasks:

- Automation of financial commitment instruments
- Application performance assurance and automated resource optimization

Operate: Continuous improvement and usage

Then evaluate performance against business objectives and look for ways to implement organizational changes and improve your FinOps practice.

Start with the fundamentals:

- Basic budgets and forecasting
- Workload planning
- Alerts and anomaly detection
- Reporting and insights

Then tackle more advanced tasks:

- Collaborative, driver-based forecasting and planning with approval workflows

IBM can help

IBM has the software, technology and the consulting expertise to help you move forward quickly with AI.

[Register](#) today for a discovery session with IBM AI experts or explore our solutions:



AI platform:

[IBM watsonx](#) provides you with the ability to train, tune and deploy genAI across your organization with trust, using your critical data wherever it resides.

AI assistant:

[IBM watsonx Code Assistant](#) helps quickly generate high-quality code and content to accelerate app modernization and IT automation efforts.

AI software:

[IBM Automation solutions](#) are designed to improve performance, manage IT spend, and simplify and optimize app management and technology operations.

AI infrastructure:

[IBM AI Infrastructure](#) offer a hybrid by design portfolio as well as advanced storage systems supporting hybrid data management for AI.

AI expertise:

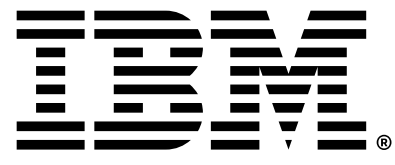
[IBM Consulting](#) has over 160,000 AI experts to help you strategize, modernize, build and manage apps with AI across your hybrid cloud environment.



IBM has long followed core principles that are grounded in commitments to trust, transparency and fairness to guide how we handle client data and insights—and how we develop and deploy new technologies.

To continue this practice in the age of AI, IBM has developed a [multidisciplinary, multidimensional approach](#) that embeds ethical principles into AI applications and processes.

With IBM's [Principles for Trust and Transparency](#) and [Pillars of Trust](#) as the foundation for our AI ethics initiatives, we're helping people and organizations adopt AI responsibly, and with clear purpose.



1. [Average Hours and Cost of Datacenter Downtime](#), IDC, July 2023.
2. [Seizing the AI and automation opportunity: The moment is now](#), IBM Institute for Business Value, 30 October 2023.
3. [CEO decision-making in the age of AI](#), IBM Institute for Business Value, 13 June 2023.
4. [The CEO's Guide to Generative AI / IT Automation](#), IBM Institute for Business Value, 10 July 2024.
5. [The CEO's guide to generative AI / Tech spend](#), IBM Institute for Business Value, 26 September 2023.
6. [The CEO's guide to generative AI](#), IBM Institute for Business Value, 18 July 2024.
7. [Gartner Forecasts Worldwide IT Spending to Grow 6.8% in 2024](#), Gartner Press Release, 17 January 2024.
8. [2024 State of the Cloud Report](#), Flexera, 2024

© Copyright IBM Corporation 2025

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
January 2025

IBM, the IBM logo, IBM Consulting, IBM Instana, watsonx, and watsonx Code Assistant are trademarks or registered trademarks of International Business Machines Corporation, in the United States and/or other countries. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on ibm.com/legal/copytrade.

Red Hat and Ansible are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

Examples presented as illustrative only. Actual results will vary based on client configurations and conditions and, therefore, generally expected results cannot be provided. It is the user's responsibility to verify the operation of any non-IBM products or programs with IBM products and programs. IBM is not responsible for non-IBM products and programs. THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

No IT system or product should be considered completely secure, and no single product, service or security measure can be completely effective in preventing improper use or access. IBM does not warrant that any systems, products or services are immune from, or will make your enterprise immune from, the malicious or illegal conduct of any party.

The client is responsible for ensuring compliance with all applicable laws and regulations. IBM does not provide legal advice nor represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.