White Paper

Five AI Building Blocks for Business Transformation

Sponsored by: Red Hat

Nancy Gohring
March 2026

IN THIS WHITE PAPER

Organizations across industries are accelerating their use of AI, moving beyond early experimentation and isolated rollouts toward broader deployment. While many enterprises have already realized productivity gains from initial AI initiatives, a growing number are now seeking more ambitious business transformations. For instance, they want AI to reshape decision-making processes, open up new revenue streams via innovative ideas, and build customer experiences that reshape how they deliver value.

The transition from orienting AI investments solely around efficiency gains toward a focus on business transformation is not trivial. It places new demands on the enterprise technology stack, operating models, and governance structures. IDC's research consistently shows that organizations are struggling to scale up the development and deployment of AI, resulting in slowed innovation, increased costs, and limited business impact.

To address these challenges, IDC recommends that organizations assess their IT environments with a view to building an AI foundation that supports the design, deployment, and operation of AI. The foundation should incorporate five critical building blocks: security, governance, data accessibility, a future-proofed technology stack, and model flexibility. Together, these elements enable enterprises to move more rapidly from ideation to production while maintaining control, compliance, and flexibility.

This White Paper outlines IDC's perspective on each of these building blocks and explains how they collectively support scalable, secure, and sustainable AI-driven business transformation.

IDC OPINION

Organizations are entering a new phase of AI adoption. The emphasis is shifting from isolated productivity improvements to broader business transformation initiatives. Today, enterprises are pursuing strategic and impactful outcomes from AI, asking questions such

as: How can AI reshape customer engagement? How can AI drive new revenue models? How can AI-powered agents identify new opportunities by safely interacting with mission-critical systems?

With these high-value goals in mind, an organization's AI foundation becomes a critical enabler, or inhibitor, of success.

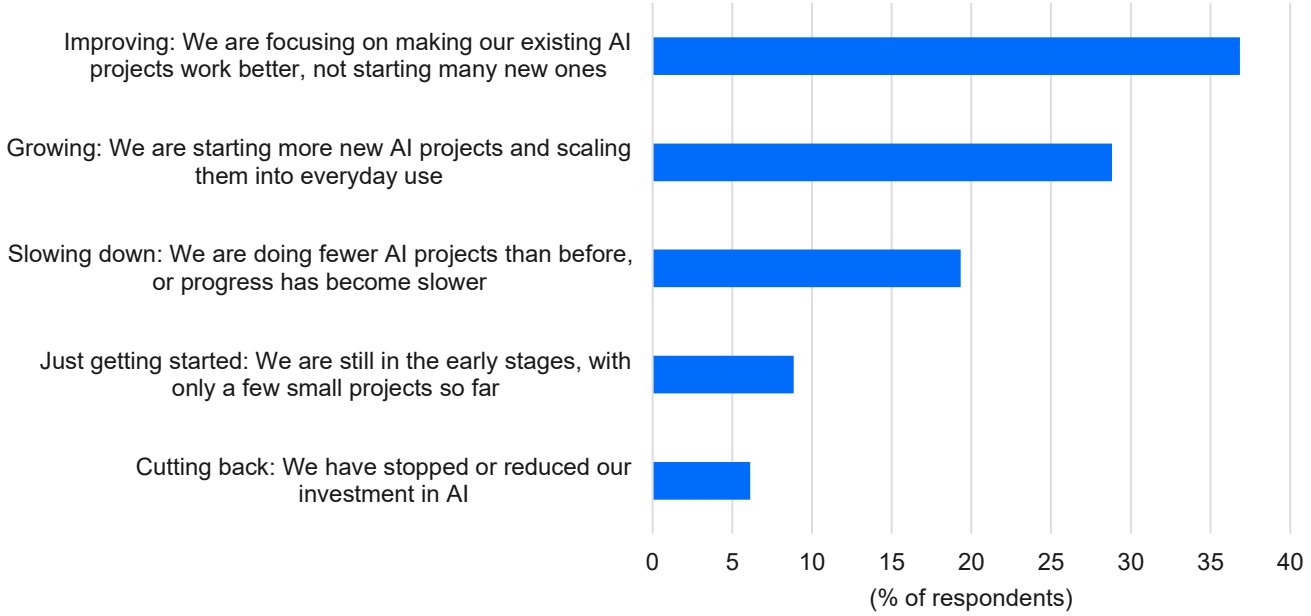
Many enterprises have spent the last couple of years experimenting with AI in silos, using disconnected tools and ad hoc processes. They often layered tools, applications, and processes on top of existing IT environments, without integrating them into broader governance and security frameworks or updating those frameworks to accommodate new challenges and threats associated with AI. In many cases, different departments were adopting competitive products in ways that increased management overhead and eliminated opportunities to break down departmental silos. While this approach may have worked well to get early pilots in place, it created friction and added complexity as organizations attempted to scale.

These challenges have led organizations to become more strategic about their AI investments. In a recent survey (see Figure 1), IDC found that over a third of respondents (36.8%) said they are focusing on improving their existing AI projects rather than starting new programs (28.8%). A significant portion (19.3%) said that they were slowing down, and a further 6.1% reported cutting back. The preponderance of organizations that said they were not growing their AI initiatives reflects that enterprises know they need to invest in improvements to address the challenges they are facing in scaling AI deployment and adoption.

FIGURE 1

Organizations' Focus: Improve Existing Projects

Q. Which of the following statements best reflects how your organization's AI program has changed in the past 12 months?



Source: IDC's *Future Enterprise Resiliency & Spending Survey, Wave 7*, September 2025 (n = 864, NA = 350, WE = 214, AP = 300)

Enterprises must build an AI foundation that supports rapid iteration while maintaining control and trust. The five building blocks outlined below reflect the core requirements for enterprise-grade AI at scale.

ENTERPRISE AI BUILDING BLOCKS

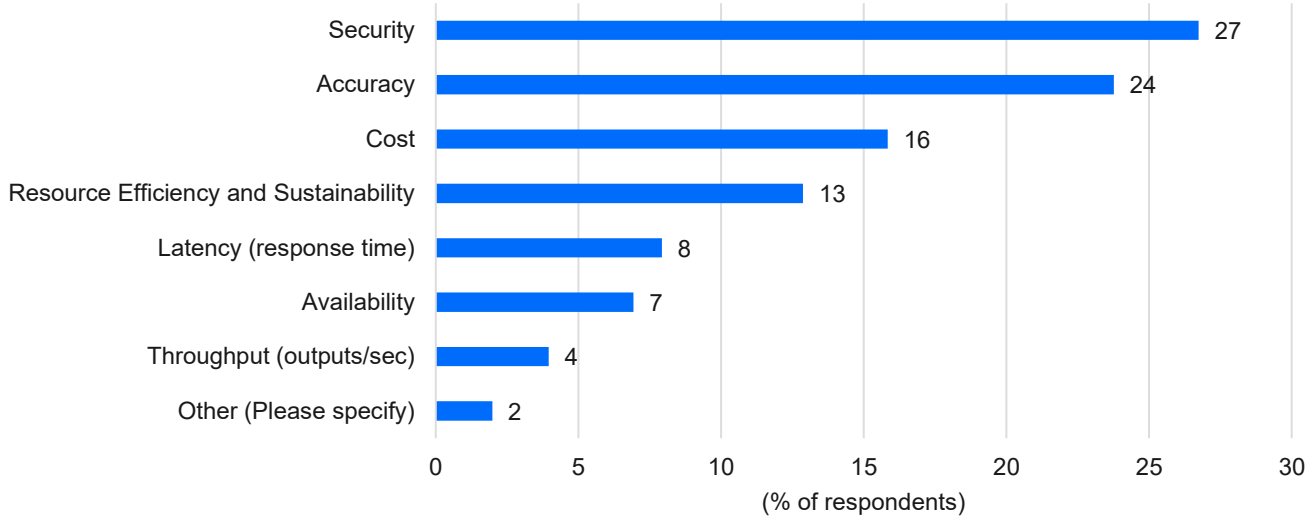
1. Security as a Core Design Principle

Security considerations must underpin and guide the building, deployment, and operation of AI applications because, as AI systems become more deeply embedded in business processes, the potential impact of security failures grows. AI introduces new attack surfaces and risk vectors, including exposure of sensitive training data, vulnerabilities in model supply chains, and misuse of AI-generated outputs. Enterprises are keenly aware of the risks. In a recent IDC study, the top GenAI inferencing challenge cited by respondents was security, a more important priority than accuracy and cost (see Figure 2).

FIGURE 2

Security and Accuracy Are More Challenging than Cost and Latency for GenAI Inferencing

Q. What is your organization's biggest GenAI inferencing challenge?



Source: IDC's *2025 US IT QuickPoll – Agentic AI and Data Q4 Survey,* November 2025

As a result, enterprises must account for security risks across the entire AI software stack, from the operating system all the way up to an agent that interacts with an end user, as well as across enterprise data pipelines.

Enterprises must not treat security for AI environments as an afterthought or bolt it on after deployment. Instead, they must integrate it into the AI infrastructure's design, tooling, and operational processes. This extends to identity and access controls, data protection mechanisms, and secure model repositories. It includes validating a trusted software supply chain and ongoing compliance monitoring, particularly for infrastructure designed to support regulated industries or use cases. As AI stacks grow more complex, achieving visibility into dependencies and validated update processes becomes even more essential. Organizations that embed security into their AI foundations are better positioned to scale deployments with confidence and meet internal and external trust requirements.

2. Expanding the Scope of AI Governance

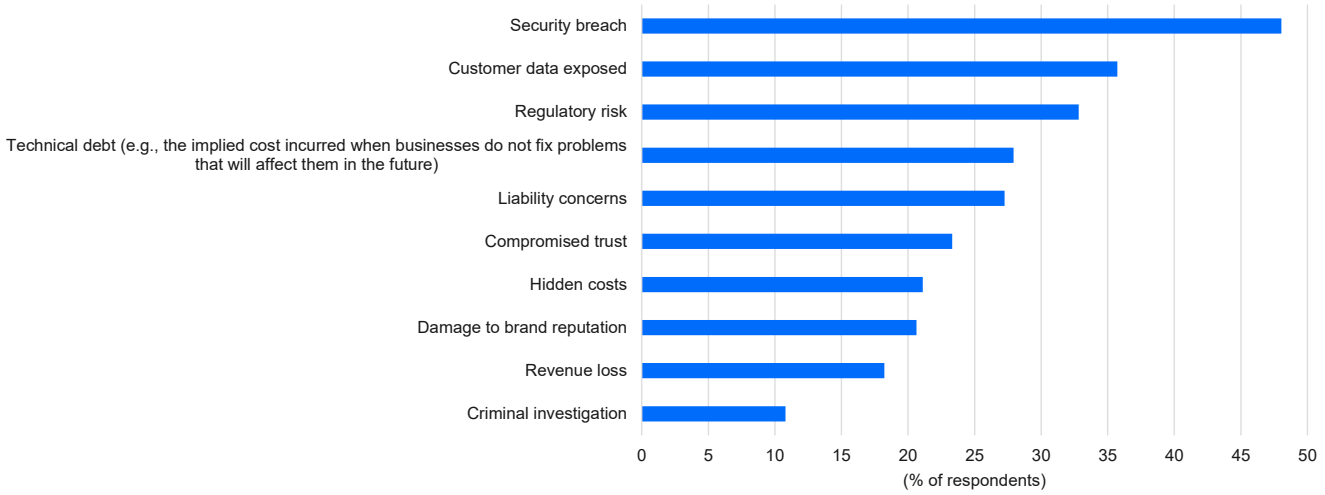
The scope of AI governance within the enterprise has expanded significantly. AI governance once focused primarily on compliance and risk management, but it now encompasses a broad range of concerns, including security, regulatory governance, internal policy adherence, use case evaluation, ROI analysis, agent discovery and tracking, AI observability, and sovereignty. However, security still tops the list. When asked about the

main concern that their responsible AI policy is intended to protect against, organizations put security breaches at the top of the list, followed by exposure of customer data (see Figure 3).

FIGURE 3

Security Breaches Are More Important than Regulatory Risk When It Comes to Responsible AI

Q. What are the main concerns your responsible AI policy is protecting against?



Note: Weighted by IT spending by country; multiple dichotomous table — total will not sum to 100%

Source: IDC's *AI Tech Buyer Survey Questionnaire Survey*, June 2025 (n = 2,276; base = respondents indicating clear principles/rules, policies, and processes/governance body currently in place at their organization)

Based on research, IDC anticipates that by 2030, AI governance will become a boardroom-level priority, receiving a level of budget and attention comparable to cybersecurity. This reflects the growing strategic importance of AI and the potential business, legal, and reputational risks associated with its misuse.

Effective AI governance should be strategic and address laws and regulations; standards, guidelines, and principles; and societal and ethical considerations. An enterprise AI governance model must extend to data, applications, platforms, and infrastructure, and it should include risk assessment, monitoring, auditability, and traceability. Transparency, accountability, and auditability are essential, particularly as AI systems become more autonomous and are deployed in regulated industries.

Organizations that invest early in comprehensive AI governance frameworks are better able to scale AI responsibly while maintaining alignment with business objectives. In contrast, governance gaps can lead to duplicated efforts, unmanaged risk, and reduced business value from AI investments.

3. Data as the Foundation of AI success

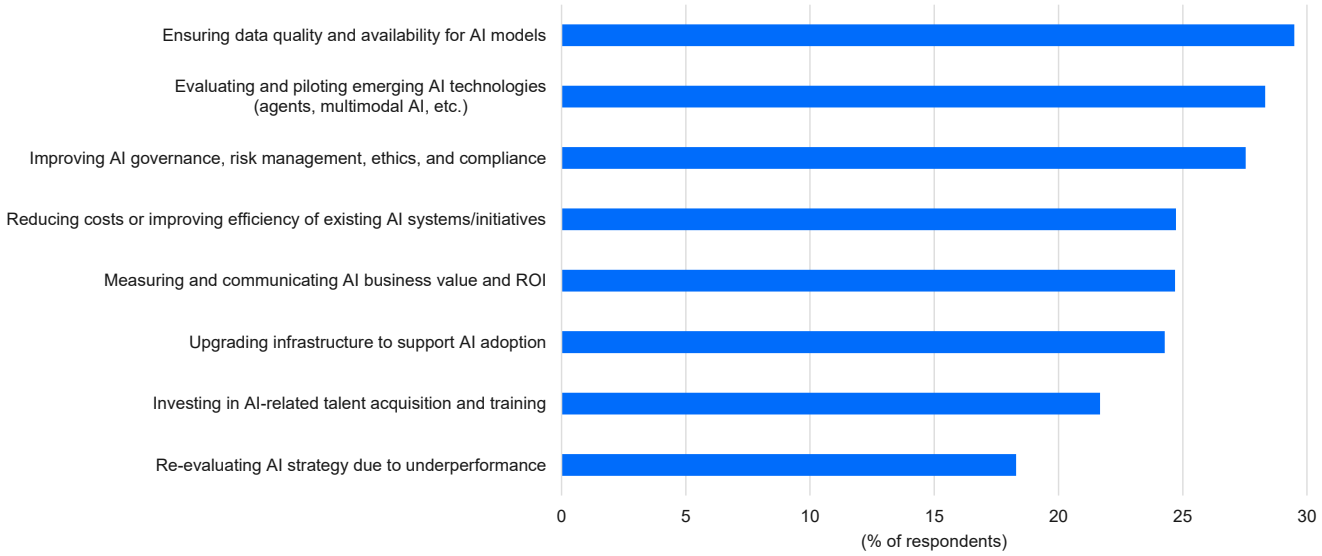
Secure access to the right data and to clean, well-managed data is critical to achieving success with AI. When corporate data is fed into AI efficiently and securely to fine-tune models or to add context via RAG pipelines or context windows, the accuracy and value of

AI outputs significantly improve. Yet data challenges remain one of the most common obstacles to scaling AI initiatives. In a recent study, IDC found that ensuring data quality and availability for AI models was the top AI investment priority for enterprises (see Figure 4). Data is often fragmented across systems, poorly governed, or inconsistently labeled, limiting its usefulness for training and inference. These issues can lead to biased outputs, reduced accuracy, and diminished trust in AI systems.

FIGURE 4

Planned AI Investments in 2026

Q. What are your organization's top 2 priorities for AI investments in the next year?



Source: IDC's *Future Enterprise Resiliency & Spending Survey, Wave 7*, September 2025 (n = 864)

Enterprises must focus on data strategies that support AI workloads, including data integration, governance, and life-cycle management. This includes ensuring that data access is secure across environments and while data is at rest, in use, and in motion. Confidential computing principles should be applied to ensure protection of data used as inputs and outputs to AI applications.

In addition, authorization becomes a critical control point, where enterprises must carefully manage permissions on a per-user basis, a complex proposition in organizations where

individuals with different permission levels are interacting with agents that may be embedded in enterprise apps or authorized to operate across applications and regions.

Organizations that treat data as a strategic asset rather than a byproduct of operations are better positioned to unlock value from AI. A strong data foundation enables faster experimentation, more reliable outcomes, and greater alignment between AI initiatives and business goals.

4. Future-Proofed Infrastructure

Many organizations did not design their existing technology stacks to support large-scale AI workloads. As a result, enterprises must assess whether their current infrastructure can meet emerging AI requirements.

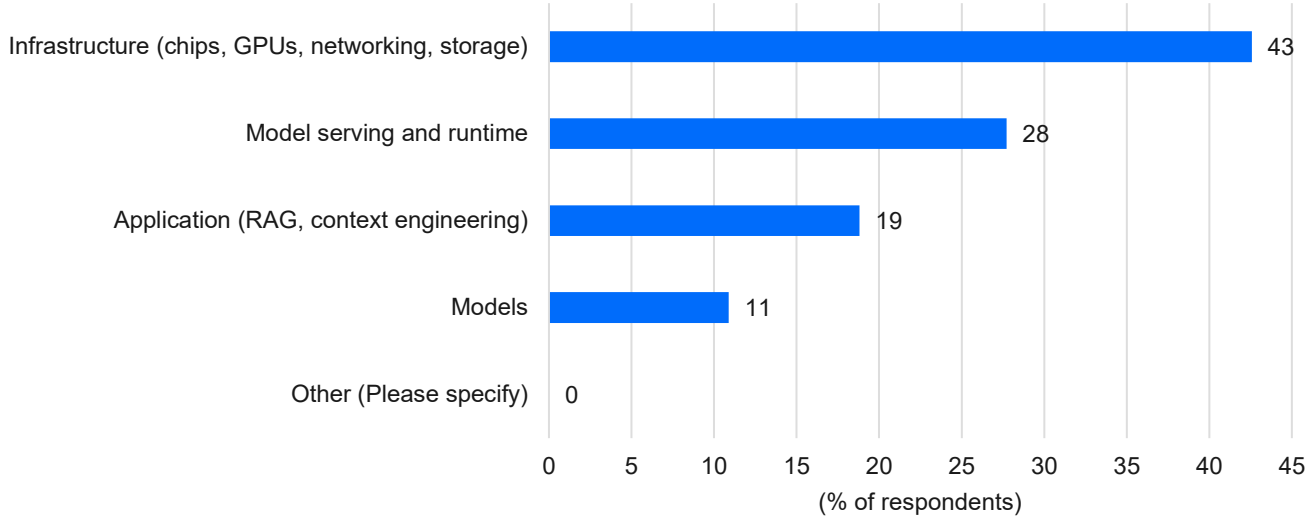
IDC recommends that organizations prioritize abstraction as a key design principle when building an AI-ready technology stack. Abstraction allows enterprises to swap technologies at different layers, from GPUs to models, as the AI ecosystem continues to evolve, without needing to significantly rebuild the software stack.

This flexibility is particularly important given the rapid pace of innovation in AI hardware. Organizations reported in a recent study (see Figure 5) that the infrastructure layer of the GenAI architecture stack is the primary way to solve their biggest inference challenges. Accelerator innovation is occurring rapidly, and building an architecture that enables the adoption of new accelerators will be key to enterprises' ability to stay competitive and match the right processor to each workload for performance and cost control.

FIGURE 5

Organizations look to the infrastructure layer to solve inferencing challenges

Q. At what layer of the generative AI architecture stack do you anticipate will be the primary area to solve your organization's biggest inferencing challenge?



Source: IDC's *US IT QuickPoll – Agentic AI and Data Q4 – Survey 2025*, November 2025 (n = 101)

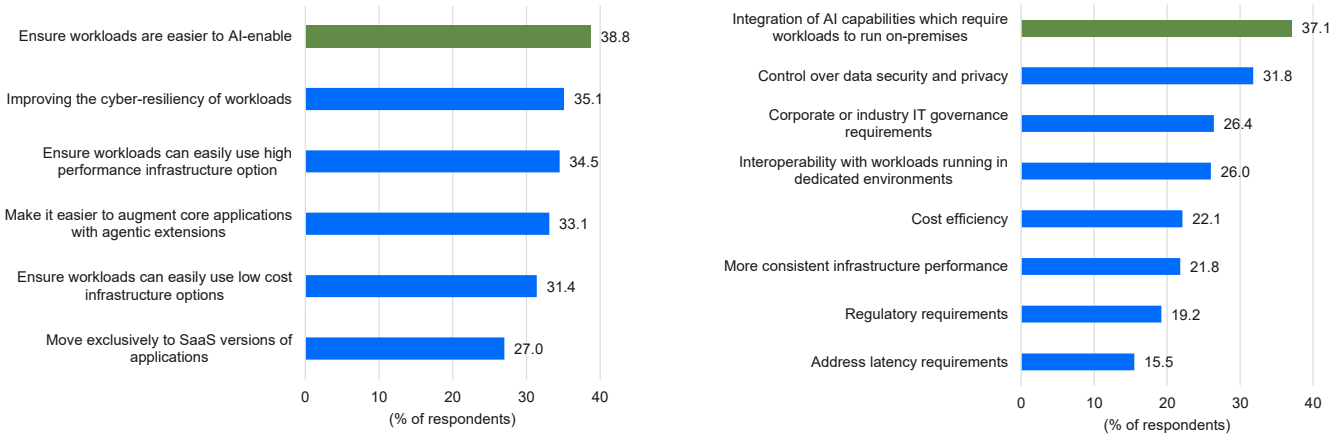
Future proofing also means designing for the ability to port workloads to different clouds or on-premises environments. Currently, AI is the primary driver for organizations that are migrating to the cloud and for those that are migrating from the cloud to on-premises infrastructure (see Figure 6). This movement in both directions reflects enterprises' need to transform their current infrastructure to meet the demands of AI. The expectation is that most organizations will adopt hybrid deployment models that combine cloud and on-premises environments. They will need to be able to move workloads easily without needing to significantly rebuild the application stack to best match an AI workload to the right infrastructure, depending on regulatory, privacy, cost, and performance requirements.

FIGURE 6

AI Is the Primary Driver for Both Shifting to Cloud and for Repatriating Workloads

Q1. You indicated your organization is prioritizing migration of applications from on premises to cloud in 2026. What are the two most important outcomes you want to achieve from your migration to cloud?

Q2. You indicated your organization is prioritizing the migration of applications from public cloud to on-premises infrastructure in 2026. What are the top two reasons for this migration?



Source: IDC's *Future Enterprise Resiliency & Spending Survey, Wave 8 & 9*, December 2025 (Q1, n = 573; Q2, n = 471)

A future-proof AI stack enables organizations to support diverse use cases without locking into a single vendor or deployment model. This adaptability helps protect investments and enables enterprises to take advantage of new capabilities as they emerge.

5. Model Selection and Customization

IDC's research indicates that enterprises expect to use an average of four models, combining proprietary, commercial, and open models to meet different requirements. This approach allows organizations to balance performance, cost, transparency, and control. However, while foundation model providers continue to advance model capabilities, adopting the newest models is not straightforward. Enterprises want to ensure that the models have been tested across multiple indicators, including performance, when running on their infrastructure of choice. Plus, switching models or adding new models to an application is not trivial.

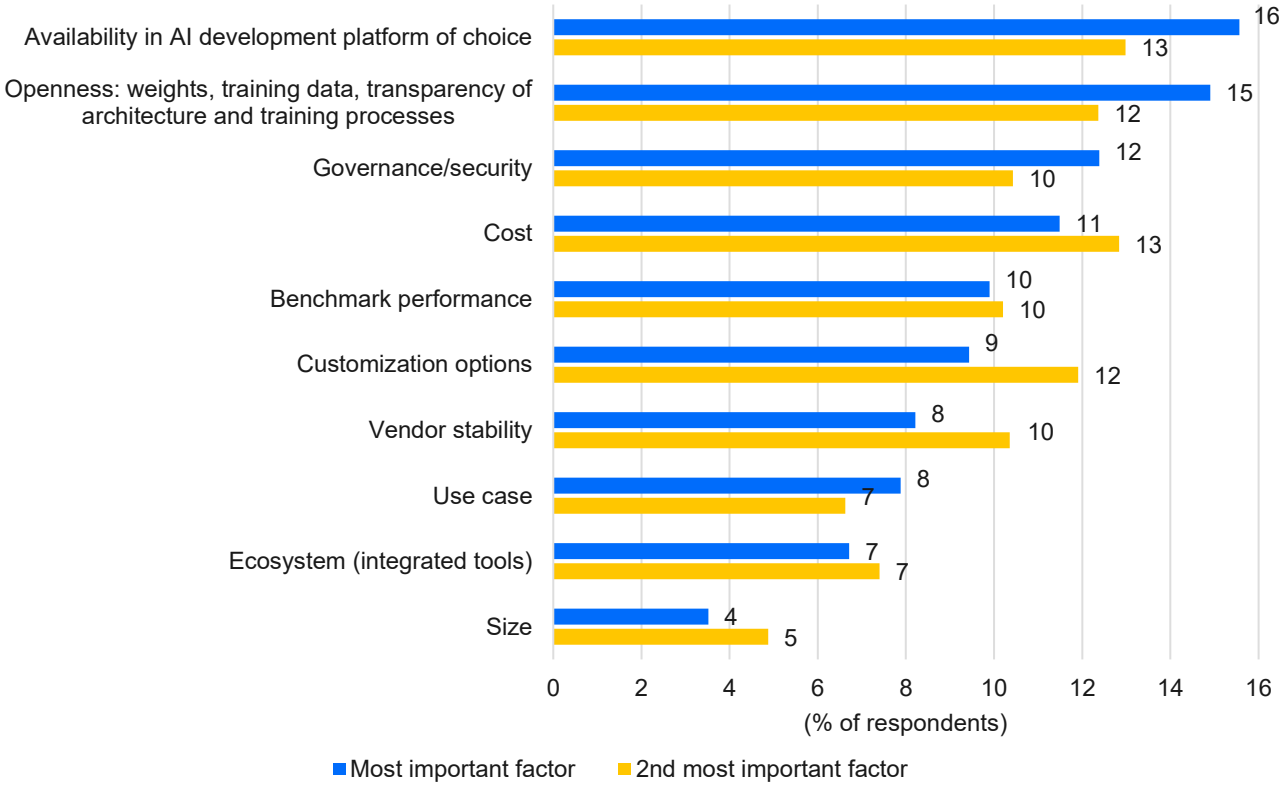
Enterprises are increasingly aware that public benchmarks are not the only factors determining model performance. Organizations have access to several techniques that can improve AI application outcomes. In fact, IDC found that availability in the AI development platform of choice is the top factor driving model selection (see Figure 7), followed by openness, indicating that enterprises are aware of and want access to techniques ranging

from fine-tuning to building retrieval-augmented generation (RAG) pipelines that impact model performance.

FIGURE 7

Importance of App Development and Governance in Model Selection

Q. What are the two most important factors in driving model selection?



Note: Data weighted by IT spending by country

Source: IDC's *Global GenAI Technology Trends Survey*, May 2025 (n = 2,280)

Effective model selection and customization require the right infrastructure, governance, and expertise. Organizations that develop these capabilities can achieve better alignment between AI outputs and business needs, improving performance and trust in AI systems.

CONSIDERING RED HAT ENTERPRISE LINUX

Red Hat Enterprise Linux (RHEL) is designed to serve as a stable, enterprise-grade operating system capable of supporting modern AI workloads that demand rapid innovation, operational control, and deployment across on-premises, cloud, edge, and sovereign environments. For technology decision-makers scaling AI into production, three areas are

particularly relevant: security architecture, hardware enablement, and operational consistency across hybrid environments.

Security Architecture

AI systems introduce expanded attack surfaces, including complex software dependencies, accelerator drivers, and sensitive dataflows. RHEL aims to address these concerns through a structured and transparent software supply chain. Red Hat built the platform under a secure development life-cycle framework, aligning it with supply chain Levels for Software Artifacts Level 3 practices, providing additional resistance to tampering and dependency risk. All packages are digitally signed, and Red Hat publishes a software bill of materials to provide visibility into included components.

RHEL also incorporates multiple cybersecurity validations and certifications, including Federal Information Processing Standard (FIPS) and Common Criteria compliance, which are particularly relevant for regulated industries. Continuous vulnerability scanning and proactive security analysis are built into the life cycle, and Red Hat backports security fixes to stable versions, enabling organizations to remediate critical common vulnerabilities and exposures (CVEs) without disrupting production systems.

At the operating level, RHEL embeds controls aligned with zero trust architecture principles. SELinux enforces application isolation through policy-driven sandboxing, aimed at reducing lateral movement risk. Secure boot and integrity attestation are designed to verify that systems start in a known, trusted state. RHEL's identity and access management tools support role-based access controls and multifactor authentication integration. Confidential computing capabilities enable data encryption not only at rest and in motion, but also in use to protect sensitive AI workloads during active processing.

Hardware Enablement

Accelerator hardware has a significant impact on AI outcomes, with the ability for enterprises to select the right accelerator for the right AI job being key to cost and performance. RHEL integrates validated GPU and AI accelerator drivers directly into supported repositories, with the goal of simplifying life-cycle management for customers and reducing reliance on unverified components. Red Hat designed its secure supply chain processes to encompass building, signing, and distributing drivers.

Red Hat is also collaborating with semiconductor partners to support new accelerator architectures as they reach general availability. Partner-optimized editions are designed to provide day 0 enablement for new platforms while remaining aligned with the main RHEL build. The aim of this is to allow enterprises to adopt performance improvements in GPUs and other accelerators without fragmenting their operating standards or compromising supportability.

Similarly, newly released AI models are made available within Red Hat's model repository. For enterprises, this validated availability shortens evaluation timelines while preserving validation and operational oversight.

Operational Consistency Across Hybrid Environments

Enterprise AI strategies increasingly span heterogeneous architectures (x86, ARM, and accelerator-enabled systems) and multiple deployment models, including public cloud, private cloud, edge, and sovereign environments. RHEL is designed to provide a consistent operating layer across these infrastructures. Its abstraction of hardware differences allows AI workloads and supporting frameworks to operate without being tightly bound to a single architecture. This consistency is intended to simplify governance, security policy application, and life-cycle management across environments.

CHALLENGES/OPPORTUNITIES

Hardware innovation cycles are shortening, and model ecosystems are evolving quickly, placing pressure on infrastructure vendors to provide near-day 1 support without compromising security, validation rigor, or life-cycle guarantees. At the same time, many organizations are experimenting with lightweight or cloud-native AI stacks that abstract away the operating system layer, potentially reducing visibility into the foundational value RHEL provides. Red Hat's ongoing challenge will be to clearly articulate and demonstrate the differentiated role of a secure, validated operating foundation in AI environments, particularly as enterprises evaluate alternative deployment models and seek to manage cost, performance, and complexity simultaneously.

RECOMMENDATIONS

Enterprises aiming to move beyond AI experimentation toward sustained business transformation should treat infrastructure decisions as strategic enablers rather than operational afterthoughts. Security, governance, and data management must be part of the AI platform's design from the outset, not layered on after deployment. Organizations should evaluate whether their current operating environments provide transparent supply chain controls, continuous vulnerability management, and support for zero trust principles, particularly as AI systems begin interacting with sensitive enterprise data and critical workflows. Embedding these controls at the foundation reduces downstream remediation costs and enables more confident scaling of AI initiatives.

At the same time, enterprises should architect for flexibility. The pace of innovation across foundation models and accelerator hardware is unlikely to slow, and technology decision-makers must be prepared to adopt new capabilities without destabilizing production environments. IDC recommends prioritizing platforms that balance the rapid enablement of

new models and hardware with validated integration, life-cycle support, and hybrid portability. Organizations that combine speed of innovation with enterprise-grade control will be in a better position to scale AI responsibly, manage costs effectively, and adapt to evolving regulatory and competitive pressures.

CONCLUSION

AI is quickly becoming foundational to enterprise strategy, but realizing its full potential requires more than access to the latest models. Organizations must establish a secure, well-governed, and adaptable technology foundation that can support rapid innovation while maintaining control over risk, cost, and compliance. By focusing on security as a design principle, expanding governance, strengthening data strategy, architecting for hybrid flexibility, and adopting a deliberate model strategy, enterprises can move from isolated AI deployments to scalable, transformation-driven outcomes. Organizations that successfully balance speed and stability will be in the strongest position to capture long-term value from AI.

ABOUT IDC

International Data Corporation (IDC) is the premier global market intelligence, data, and events provider for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries. IDC's analysis and insight help IT professionals, business executives, and the investment community make fact-based technology decisions and achieve their key business objectives.

Global headquarters

One Beacon Street
Suite 33100
Boston, MA 02108
USA
508.872.8200
X: @IDC
blogs.idc.com
www.idc.com

Copyright notice

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2026 IDC. Reproduction without written permission is completely forbidden.