ANALYST BRIEF
Sponsored by: Composable

The immense potential of generative AI has driven organizational urgency and investment. Yet organizations are realizing that these solutions are complex and multifaceted. This Analyst Brief looks at the role of LLM software platforms to improve how technology buyers build, deploy, manage, optimize, and scale GenAI.

Moving Beyond the Model to Maximize the Effectiveness and Reach of Generative AI Deployments

July 2024

Written by: Matt Arcaro, Research Director, Computer Vision and AI

Introduction

The catalyst for today's generative AI (GenAI) revolution began when the world saw the potential of the LLM that underpinned OpenAI's ChatGPT service (built on a fine-tuned version of its GPT-3.5 series foundation model) in November 2022. Although monumental from a software packaging and an applied technology perspective, ChatGPT reflected the culmination of research and investment advancements across academic and industry channels to extend LLMs' capabilities, accuracy, coherence, efficiency, and use more broadly. Even more impressive, in the approximately 18 months since the launch of ChatGPT, 89.1% of worldwide technology decision-makers in IDC's April 2024 *Future Enterprise Resiliency and Spending Survey, Wave 3,* indicated that they already investigated, experimented, prototyped, and/or deployed GenAI within their environments. This urgency and pace for GenAI adoption and pursuit is largely unparalleled in the technology ecosystem and underpins the immense value and broad applicability that organizations expect the technology to unlock.

AT A GLANCE

KEY TAKEAWAYS

- 89.1% of worldwide technology decision-makers indicated that they already investigated, experimented, prototyped, and/or deployed GenAI within their environments.
- 70.2% of worldwide technology decision-makers indicated that they believe GenAI has already disrupted or will significantly impact their business over the next 18 months.
- Organizations adopting LLM software platforms provide them a considerable advantage over their peers bringing GenAI solutions and applications to market.
- These advantages go beyond solution deployment and extend beyond life-cycle management and continuous improvement.

Yet the speed of GenAI adoption doesn't tell the whole story. Organizations have been working to build the airplane while "in the air," forcing them to balance, course correct, and respond as critical factors, including technological advancements, organizational requirements, processes, and use case prioritization, continue to evolve. This idea of working smarter and not harder is rapidly becoming a critical adage separating leading and lagging organizations, especially as 48.9% of these same technology decision-makers indicated that they have already engaged in more than 25 GenAI proofs of concept.

So what is the secret of leading organizations to drive greater success with GenAI? IDC posits that one key theme that drives leading organizations' successful adoption of GenAI comes from not focusing on the underlying model. These leading organizations understand that the model ecosystem remains extremely dynamic and variable, making it nearly impossible to focus on a single model. For example, IDC sees some organizations struggling with GenAI focusing on a single model and investing considerable time and effort to fine-tune it for their organization, a given project, and/or a single use case. In these scenarios, these model fine-tuning efforts may improve the effectiveness of the output/inference of an LLM, but they do so at the expense of being able to provide sufficient time and effort to understand the end-to-end solution requirements needed for the ultimate production system. On the flip side, IDC often sees organizations doing the opposite: struggling with sourcing and integrating multiple LLM models into their process flows. They often correctly understand the need to embrace model diversity and innovation without realizing how to abstract the structural differences, nuances, and performance differences that a multiple model–based approach requires. Further, these organizations rely on customized, brittle, and labor-intensive frameworks and practices to unify and normalize the flow across multiple models and providers. This approach may work today but will be unable to survive the test of time.

IDC recommends that organizations look to the expanding ecosystem of LLM software platforms to abstract away the complexity of GenAI. These platforms enable organizations to focus on what they do best while helping accelerate and future proof how they experiment, build, deploy, manage, and scale GenAI more broadly.

Benefits of LLM Software Platforms

GenAI is not fading away. About 70.2% of worldwide technology decision-makers in IDC's April 2024 *Future Enterprise Resiliency and Spending Survey, Wave 3,* indicated that they believe GenAI has already disrupted or will significantly impact their business over the next 18 months. These organizations, including both technology and line-of-business leaders, understand that effectively harnessing GenAI is a skill set essential to unlocking future growth and differentiation. Thus organizations must quickly identify, procure, and deploy technologies that improve the efficacy, speed, repeatability, control, cost, and use of GenAI. LLM software platforms remain an essential puzzle piece to help organizations think beyond just the model, including helping deliver:

- **Expanded GenAI use case and user reach:** Technology organizations face challenges in managing the demands and expectations that GenAI places on their teams. LLM software platforms enable these organizations to streamline the GenAI development process, allowing them to move more quickly, build on successes, correct failures, reuse results, and broaden how and where GenAI can be implemented. Further, LLM software platforms allow organizations to target multiple use cases in parallel, reducing the need to serialize each use case individually while providing a seamless UI/UX that reduces the need for users to understand (or care about) the nuances of today's disjointed AI ecosystem's tooling.

- **Embedded life-cycle monitoring and management capabilities:** This phenomenon around under-indexing on post-deployment monitoring, management, and continual improvement is nothing new in the AI world. Unfortunately, organizations spend so much time building, testing, validating, and deploying their AI-powered solutions that anything else becomes an afterthought. LLM software platforms give organizations vital visibility into their solutions, offering real-time performance insights, recommended optimizations, and the ability to implement updates remotely.

- **Comprehensive data privacy and security compliance, auditing, and data controls:** As organizations move through the continuum of deploying GenAI across more and more use cases, they will need to integrate and utilize

critical, often sensitive, data sources. LLM software platforms have designed sufficient guard rails, data controls, and auditing capabilities to ensure an organization can limit and manage how and where their data is being used, maintain adequate access control, and perform version control.

- **Improved solution accuracy and coherence across varying models and provider approaches:** As outlined in the Introduction section, organizations pursuing GenAI often follow the pathway of going deep into (i.e., fine-tuning) a single model or utilizing a multimodel and/or provider approach. LLM software platforms help organizations with either approach. From streamlining the creation of training data sets, benchmarking model performance, normalizing prompt formats across models, creating complex orchestration pipelines, seamlessly transitioning to new foundation models, and monitoring solution performance to helping with cost and access controls, LLM software platforms are outcome oriented to help organizations build GenAI applications that maximize their business requirements and objectives.
- **Tooling to build complex workflows that include the extension and integration of third-party data sources:** GenAI applications are only as effective as their ability to access and integrate sufficient relevant contextual information and data. LLM software platforms can unify and optimize the relationship between structured and unstructured data sources and LLMs across various documents, chatbots, and emerging assistant/agent use cases. These platforms include capabilities that help with indexing/vectorization, retrieval-augmented generation, and usage within a workflow pipeline.
- **Seamless future proofing to accelerate the incorporation of new models and capabilities:** Over the past 18 months, it has become evident that each model generation will improve its accuracy, efficiency, coherence, and applicability to additional use cases. Organizations building GenAI models and applications must leverage these advancements without considerable refactoring, retraining, or constant reconfiguration. LLM software platforms understand they are designed to do just that, specifically ensuring that organizations and customers have the flexibility to adapt to new and emerging capabilities and models in a lightweight, often zero-touch way.

Considerations When Evaluating LLM Software Platform Providers

Organizations researching, experimenting with, or pursuing LLM software platforms need to recognize that the nascency of the GenAI technology ecosystem and pace of innovation create a platform ecosystem that spans a wide range of capabilities, business models, and product offerings. In this way, not all platforms are created equal. IDC has identified several strategic and technological attributes that organizations should consider when pursuing an LLM software platform. These attributes include:

- **Broad ecosystem technology support, including integrating multiple open source and proprietary LLMs and providers:** The portfolio and variability of open source and proprietary LLM models and providers and their approaches remain extremely diverse. This diversity is driving an increasing wave of provider differentiation, including how and where they look to invest in their future R&D. Organizations should consider LLM platform providers that make openness a key strategic tenet to enable the ability to build today and support a pivot tomorrow (as necessary).
- **Highly tunable model orchestration strategies, including support for synthetic LLMs:** As organizations look to deploy GenAI more broadly, complex model orchestration and distribution strategies built on logic and business rules, including the integration of multiple LLMs, will be required to deliver performance that aligns with stringent

business requirements. Organizations should consider providers that develop configuration tools designed to maximize performance across various parameters and configurations.

- **Sandboxing capabilities that enable the ability to experiment, validate, and productize LLM-enabled solutions and applications quickly:** Building GenAI solutions and applications is a very iterative and (often) experimental process for organizations. Organizations must prioritize LLM software platform providers that provide them with the environment and tooling to quickly build initial prototypes, understand performance, iterate based on feedback, and progress the solution toward production.
- **Multi-persona (i.e., business analyst, prompt engineer, developer, product owner) tooling to enable collaboration across user bases:** Organizations need to democratize the effective use of GenAI beyond just ML engineers and data scientists. Although these team members are important for highly technical, low-level AI tasks and data analysis, organizations that broaden platform access to a greater range of workers will extend GenAI's reach and the resulting business value. Prioritize LLM software platforms that thoughtfully integrate purpose-built support (and technology abstraction) to enable multiple user personas and build LLM applications and solutions.
- **End-to-end governance, including support for auditability, monitoring, cost, and change management:** IDC cannot understate the importance of LLM solution life-cycle management. Organizations mistake solution deployment for the finish line, when realistically, this is just the starting point. They need to prioritize LLM platform providers that offer comprehensive and proactive support for the entire life cycle, including auditability, monitoring, cost, and change management.
- **Progressive technology development strategy that includes a well-thought-out road map of features:** No two organizations have the same technology stack, software architecture, approach to procuring technology, vendor relationship, future development road map, and/or GenAI ambitions. This variability forces organizations to prioritize vendors with strong alignment and support based on their current environments while having a cohesive strategy and execution framework to progress an organization's GenAI maturity.

Understanding How Organizations Are Evaluating LLM Software Platform Providers

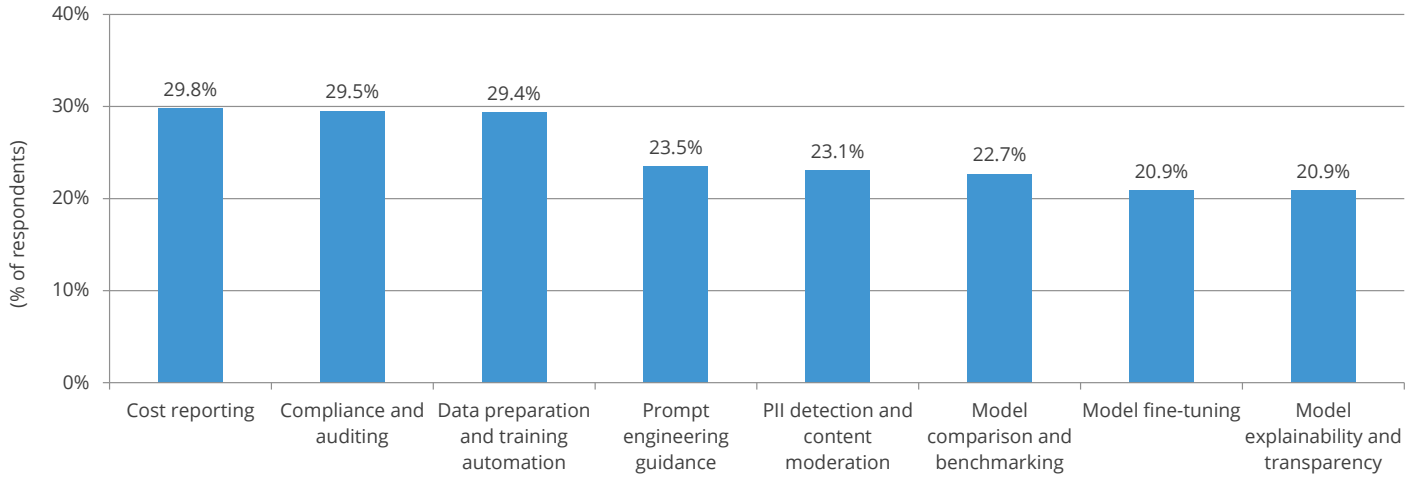
Figure 1 provides insight into how organizations view and evaluate LLM software platform providers. Such a question provides several key takeaways that reinforce the market need for LLM software platforms and evaluation considerations:

- **Respondents indicate that there isn't a universal framework or single capability that drives their LLM software platform evaluations:** This is to be expected in today's frenetic GenAI technology environment. Further, IDC sees this as an opportunity for organizations to build a coalition of stakeholders across different functions to outline, affirm, and quantify the most critical capabilities. These exercises ensure that diverse voices are considered and better reflect how GenAI is most successful when technology and line-of-business stakeholders align.
- **Cost reporting was viewed as the most critical LLM software platform capability:** This is interesting and noteworthy because cost reporting, although technically a capability, is just a configurable output. Subliminally, this says that LLM software platform providers need to consider, represent, and thus configure cost throughout the end-to-end solution life cycle.

- **Compliance reporting and auditing was a close second:** Respondents indicate that processes, guardrails, and tooling are paramount to ensure that GenAI solutions and applications adhere to process and compliance requirements. This important step reinforces that organizations will progress cautiously to ensure data privacy and protections are in place and verifiable.
- **GenAI pipeline capabilities vary in importance:** This response makes sense, as organizations vary differently in their experience, ambitions, and success with GenAI.

FIGURE 1: ***How Are Technology Buyers Evaluating LLM Software Providers?***

Q Which of the following capabilities were most important to consider in evaluating LLM software platform providers? (Please select all that apply.)



n = 881

Source: IDC's Future Enterprise Resiliency and Spending Survey, Wave 1, February 2024

Conclusion

Organizations have realized that the potential value of GenAI is too big to ignore. They have been scrambling to adapt their people, technology resources, and processes to maximize the effectiveness of GenAI. Yet many of these efforts to date have not been entirely fruitful. A big driver for these inefficiencies comes from an undue focus and investment in the model or underlying LLM rather than a pervasive view of the end-to-end solution. As a result, IDC sees leading organizations looking to LLM software platforms to help them maximize the effectiveness of their GenAI development efforts, including enabling performance improvements, solution flexibility, cost management, third-party data usage and integration, and life-cycle management capabilities.

About the Analyst

Matt Arcaro, *Research Director, Computer Vision and AI*

Mr. Arcaro spearheads IDC's research and thought leadership initiatives for computer vision (CV) AI tools and technology. In this role, Mr. Arcaro comprehensively tracks the software tools, technologies, and ecosystem trends that are catalyzing the expansion of CV more broadly across business and consumer use cases.

MESSAGE FROM THE SPONSOR

Composable is the only API-first platform for building intelligent applications and services using LLMs. With Composable, enterprise teams can harness AI by leveraging multiple inference providers and models to create LLM-powered tasks that safely automate and augment their business, without the burden and complexity of managing disparate APIs, security models, or prompt formats.

Composable is more than an LLM application development framework. The end-to-end platform provides governance, fine-grained security and orchestration, while supporting a large number of inference providers (Google, OpenAI, Amazon, Replicate, Hugging Face, TogetherAI, etc.) and models (GPTx, Claude, Cohere, AI21 Labs, Titan, Llama, Mistral, and many others), and brings a new virtualization approach to operationalize them.

Learn more about the future of LLM operations at becomposable.com.

IDC Custom Solutions

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

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