Quick Answer: Which Cloud Provider Offers the Best Generative AI Business Outcomes?

Published 26 July 2023 - ID G00793924 - 5 min read

Jim Hare

Initiatives: Executive Leadership: Cloud for Business Enablement; Artificial Intelligence; I&O Platforms; Infrastructure for Technical Professionals

Many executive leaders expect generative artificial intelligence to transform their businesses. Some want to align their choice of a strategic cloud provider with the anticipated future market leaders in GAI; however, the market is still too nascent to predict winners and losers.

Quick Answer

Should I switch strategic cloud providers in anticipation of the future adoption of generative artificial intelligence (AI) services?

- Generative AI (GAI) is an emerging collection of technologies and practices. Although there will be few large-scale foundational models (FMs), there will be a massive ecosystem of applications and tools that build on, extend and commercialize the underlying FMs.
- Although hyperscale, cloud. integrated infrastructure as a service and platform as a service (IaaS+PaaS) providers offer a range of GAI capabilities, it's too early to predict which will emerge as market leaders in any of the capabilities categories. Therefore, it's also too early to evaluate GAI capabilities as part of your selection of a long-term strategic cloud provider for IaaS and PaaS.
- If you have a specific GAI project, default to your existing strategic cloud provider, if they have the capabilities that you need. Otherwise, choose vendors with the capabilities that meet your project requirements in the context of your multicloud workload placement policies.
- If you plan to build your own GAI solution, your training data would ideally reside alongside the infrastructure that you will use to train or fine-tune your model. However, this isn't necessary, because the datasets are typically small; therefore, the associated data transfer costs are likely to be modest.

More Detail

Executive leaders in organizations that haven't yet chosen a strategic cloud provider (e.g., Amazon Web Services [AWS], Microsoft Azure or Google Cloud Platform [GCP]) may wonder how much GAI should influence their multicloud strategy and provider selection. Executive leaders in organizations that have already chosen a strategic cloud provider are increasingly concerned that technology developments in the field of GAI may radically shift the market dynamics for cloud IaaS+PaaS. They also worry that they may need to switch providers to effectively leverage GAI.

At this stage of the market, these concerns are misplaced. GAI is evolving rapidly. It is impossible to predict which cloud providers will be best-positioned to deliver the optimal business outcomes for a particular business. The provider best-suited to immediate GAI needs may not be the best provider for long-term GAI needs. Gartner expects organizations will successfully integrate with GAI solutions across hybrid and multicloud environments.

Most customers will continue to use FMs, rather than trying to train an FM from scratch (a costly undertaking often involving massive datasets). Providers vary in data science (DS) and machine learning (ML) capabilities, as well as support for different FMs. You may reasonably choose to run GAI pilot projects on a tactical cloud provider of interest, regardless of your strategic cloud provider choice (or preference for on-premises infrastructure). If you need to move a dataset into or out of a cloud provider, this can be done at a reasonable cost. Typically, the training dataset necessary to tune a model is modest in size. Good results are often achievable with just a few dozen to a few hundred training examples. This represents a trivial data transfer cost.

Although applications such as ChatGPT that run on top of large language models (LLMs) are seizing the public imagination, conversational AI is only one possible application of GAI. In turn, it is only one of many forms of AI. Furthermore, executive leaders should be careful to distinguish cloud provider strengths and weaknesses in:

- GAI capabilities embedded in SaaS — such as in Microsoft 365 and Google Workspace
- GAI capabilities accessed through API-based SaaS — such as translation and computer vision APIs
- Developer productivity tools with embedded GAI capabilities — such as Amazon CodeWhisperer and Microsoft Copilot

- PaaS-based, cloud-based AI developer services (CAIDS) that can assist organizations in developing GAI capabilities — for example, Magic Quadrant for Cloud AI Developer Services
- Cloud IaaS resources that can be used to power training and inference workloads

You should expect that hyperscale IaaS+PaaS cloud providers will participate in all aspects of GAI, including infrastructure, platforms, tools and applications. Existing services that support AI/ML training and inference can be used by customers that want to build their own models or tune existing models. Over time, they will also be able to:

- Embed GAI into their existing services
- Support open-source and commercial FMs (including their own FMs) that enable customers to build their own GAI solutions
- Deliver solutions that will make it easier for customers to use GAI
- Deliver new solutions enabled by GAI capabilities
- Provide capabilities to assist customers in governing and securing GAI solutions
- Invest in hardware solutions that accelerate training or inference — such as AWS's Trainium and Inferentia chips, or Google's TPUs
- Use GAI to support their business and technical operations
- Attract partners and participate in emerging GAI ecosystems, including industry-specific ecosystems

Although hyperscale cloud providers are likely to deliver similar DS and ML core capabilities, each is likely to differentiate by hardware strategy (relationships with third-party vendors such as NVIDIA and first-party proprietary hardware) and access to proprietary data used to drive industry-specific solutions.

Develop a strategy for GAI and integrate and harmonize it with your enterprise's overall AI strategy, focusing on use cases for applying the technology. Because GAI technology is evolving so quickly, avoid the temptation to go forward alone. Find support and knowledge from partners and third-party organizations operating in this space.

There is no need to panic and radically rethink your cloud provider choices, especially if you are already using a market leader. Strive to understand the GAI strategy and roadmap of your strategic providers, but recognize that these may evolve rapidly and significantly. Go forward with your existing multicloud strategy (or single-cloud strategy). However, you may want to direct your cloud governance team — such as your cloud center of excellence (CCOE) — to revise your multicloud workload placement policy so it provides prescriptive guidance for GAI use cases.

Recommended by the Authors

Board Brief on Generative AI

Use Generative AI in Applied Innovation to Drive Business Value

Applying AI — A Framework for the Enterprise

A Multicloud Strategy Is Complex and Costly, but Improves Flexibility

How to Design a Multicloud Workload Placement Policy

Solution Path for Assessing and Selecting Public Cloud IaaS and PaaS Providers

Magic Quadrant for Cloud Infrastructure and Platform Services

Magic Quadrant for Cloud AI Developer Services

Quick Answer: How Should Organizations Prepare for the Addition of Generative AI to the Microsoft Stack?