

Generative AI is Here: Are you ready?

Best practices for bringing
AI to organizational data

Table of Contents

Building the foundation for success	1
The vital role of your data in Generative AI.....	2
Achieving Generative AI success through organizational readiness	3
Considerations for organizational readiness	4
Leveraging the right data requires collaboration.....	5
Keep Generative AI models close to your data	6
Generative AI and data management	7
Right-sizing your generative AI models.....	8
AI deployment models: Assessing cost and value tradeoffs.....	9
Having the right infrastructure for your models	10
Bringing AI to your teams.....	11
The last mile	12
New beginnings	13
Getting started with Generative AI	14
Dell Technologies can help accelerate your journey.....	15



Building the foundation for success

Generative AI has captured everyone's attention

While still in its infancy, Generative AI is already being leveraged on a massive scale by organizations and individuals. In a Dell survey, 91% of business and IT leaders report they are using Generative AI in their lives in some capacity, and 71% indicate they've used it for work¹.

Consumer-grade AI (think: ChatGPT and DALL-E) has already impacted society in innumerable ways, revolutionizing how people interact with technology and creating human-like outputs of conversational text, images, audio, video, and software code.

Where to start

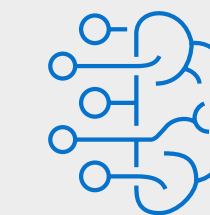


Risks and rewards

In turn, Generative AI will be equally impactful to organizations, transforming them in ways that visionaries are only on the cusp of imagining. In the rush toward adoption, however, it's tempting for business leaders to move too quickly and bypass important implications involving their data, compliance, governance, and other risks.

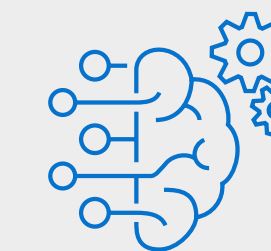
Organizations have different considerations than individuals as they embark on their Generative AI journey. They cannot risk customer trust and the high value of their data for the reward of being first to the finish line.

Generative AI isn't perfect, but your organization *needs* it to be. This eBook explores how you can proceed both with caution and confidence in your Generative AI endeavors.



AI

Artificial intelligence (AI) applies advanced analysis and logic-based techniques to interpret events, and support and automate decisions and actions.²



Generative AI

A set of technologies and techniques that leverage massive corpuses of data, including large language models, to generate new content (e.g., text, video, images, audio, code). Inputs may be natural language prompts or other non-code and non-traditional inputs.³

1. Dell internal research

2. [Gartner Information Technology Glossary: Artificial Intelligence](#)

3. [Forrester: Ignoring Generative AI Will Be A Costly Mistake for Enterprises](#). Forrester. February 14, 2023.

The vital role of your data in Generative AI

Data is the differentiator

While Generative AI has already shown significant potential to transform organizations and industries, business and IT leaders are just beginning to consider the breadth of possibilities for their organizations.

One thing is certain, however: Quality data is the critical element of your Generative AI projects, and the more you use your own data for your projects, the higher value your projects will be to your organization. The potential when you leverage your own data is endless.

As you begin your Generative AI journey, it's important to develop a comprehensive strategy to harness the uniqueness of your data, resulting in specialized projects that add value to your organization. Choosing the right projects involves careful evaluation of the requirements, risks, demands, inputs, and outputs for each use case. As your Generative AI capabilities increase, so will your ability to fine-tune your models, enabling more innovation throughout the enterprise. Establishing this early vision around the value of your data will help prepare your organization for the future of work.

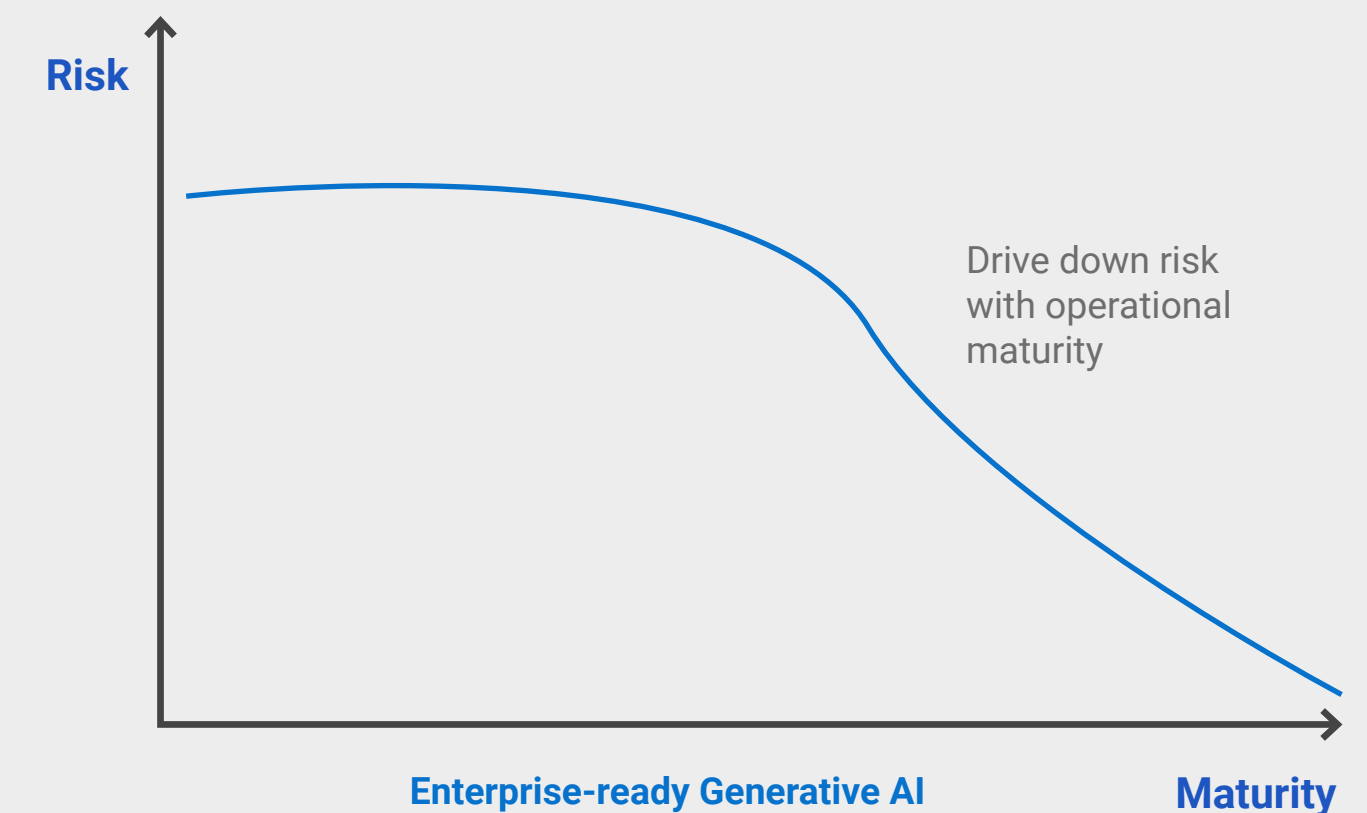
Understanding data risks in Generative AI

Data and risk go hand-in-hand. Data will drive your Generative AI projects forward, but you also need to assess the potential risks of hosting Generative AI models in public clouds, including:



Managing risk and enhancing value

As you begin your journey, you need to make corresponding investments in technology and training to reduce risk and deliver more control and rising value to your organization.



Achieving Generative AI success through organizational readiness

Every organization has unique Generative AI needs

Generative AI technology can benefit organizations in innumerable ways. While enterprises have unearthed hundreds of use cases across every vertical, landing on the right use cases for your organization, based on your own data and unique needs, will help you achieve success more quickly.

How to begin



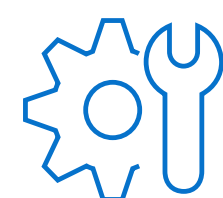
People and teams

Prepare your organization to address the Generative AI opportunity. IT organizations look inward, while the business looks outward.



Processes and policies

Define and communicate how your organization will leverage AI. It's critical to get employees engaged with AI as an aspect of your business.



Technology

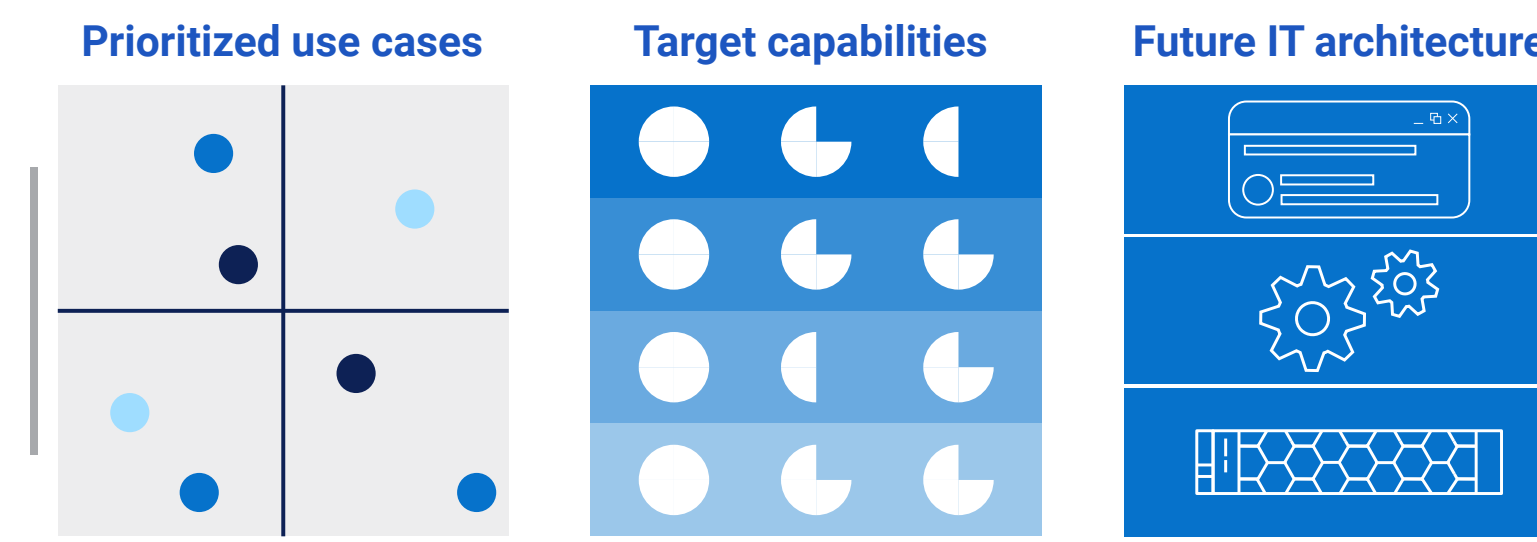
Deliver secure access to Generative AI across your organization. Have an open discussion with business stakeholders on how IT will provide these capabilities.

Define what your “as-is” and “to-be” states are

Generative AI use cases will drive technical infrastructure planning and deployment. Here are some points to help determine your level of investment:

Capture the “as-is” current state of your environment to develop topology diagrams and document information on key technical systems. Using these insights, you can determine the strategic vision and guiding principles for your future Generative AI projects.

Understand planned investments, long-term strategy decisions, and organizational and financial information for the targeted environment. Build a proposed “to-be” future state blueprint and solicit feedback from all stakeholders before finalizing your vision.



With the explosion of AI-driven automation, specialists can spend less time analyzing images and quantifying data for simpler, repetitive tasks. Not only does this provide specialists with more time for complex issues, but they can deliver more accurate diagnoses faster.⁴

—Nikolas Stathonikos

Principal Investigator, AI Development and Implementation, UMC Utrecht

Considerations for organizational readiness

If you want to build out a highly specialized capability in your industry, you'll likely need to train your own model, and the technical requirements to do so will be high. Conversely, standing up private instances of pre-trained Generative AI large language models (LLMs), such as LLaMA 2 or Falcon, offer advantages of speed and deployment, but may require fine-tuning if they do not meet all your requirements. Retrieval-augmented generation (RAG) allows you to incorporate enterprise-specific

knowledge and is comparatively simple to deploy and use. Organizations can also purchase SaaS that's Generative AI enabled.

Whichever path you take, when determining whether you should build or buy your Generative AI solution, you must take into account your required business outcomes, the applicability of your organizational data, and your potential use cases.



Build

Industry specific

- Drug discovery
- Material design
- Optimized crop yields

Business function

- Marketing
- Sales
- Procurement

Productivity

- Salesforce
- Microsoft 365
- Box

Buy



Data governance

Generative AI projects require strict compliance with data and privacy regulations, including data sovereignty restrictions. While it's highly beneficial to leverage your data for Generative AI projects, stringent data handling is necessary.

Preparing your data for AI

Your organizational data is best suited for Generative AI when it's cleansed, consistent, and centrally stored. Be sure to monitor and augment your data regularly to ensure reliability, consistency, and integrity.

Proof-of-concepts and minimum viable product

Demonstrate the feasibility of Generative AI in your organization with POC projects. Small, easily attainable projects will attract the attention of employees and leaders and increase the perception and value of Generative AI across the organization.

AI as a long-term investment

While it's best to start small, it's also important to realize that AI is here to stay and will become a vital cog in operations. We are still at the beginning stages of the AI revolution, and organizational budgets must account for significant future investments.

Leveraging the right data requires collaboration

Mapping use cases to your data

The best Generative AI use cases leverage your rich data and involve the right blend of skill sets, budgets, and resources to support them. It's important for business and IT teams to collaborate to establish priorities. One point is certain: Business stakeholders will likely want to move faster than IT. This means you need to find the right balance between speed and risk.

When considering your internal readiness, understand that your data is best suited for Generative AI when it's cleansed, consistent, and centrally stored. Be sure to monitor and augment your data regularly to ensure reliability, consistency, and integrity. Also, Generative AI projects require strict compliance with data and privacy regulations, including data sovereignty restrictions.



Many stakeholders are interested in Generative AI

Collaboration between business and IT teams is critical to establishing use-case priorities and arriving at a consensus on readiness.

Key personas and responsibilities:

C-suite and business leaders: Identify business opportunities and priorities that can be solved with Generative AI.

IT leaders: Ensure the right level of technical infrastructure to support current and future projects, including reliability, scalability, and security. They are co-collaborators in data decisions.

Data scientists: Design and deploy foundational Generative AI models.

Domain experts: Deliver specific domain knowledge and data for training and validating Generative AI models.

Keep Generative AI models close to your data

Why data proximity matters

In terms of workload placement, to get the best outcomes, put it in the environment that makes the most sense based on your business requirements and technical needs. However, keep in mind that with Generative AI, a lot of data that didn't have a strong use case previously becomes interesting and may prompt you to rethink placement.

Following are characteristics that can help determine the most appropriate environments. Your own data implementations and preferences may be different, but these concepts, also conveyed in the diagram, should help you get started.

Considerations for Generative AI placement

Secure access: The simplest approach often means maintaining complete control over your own infrastructure, data, and services.

General purpose use cases: Public cloud has an advantage for general use cases, but for specific use cases, private environments allow more customization.

Accuracy and customization: Private cloud allows more customization to access timely data and create guardrails for more accuracy and better model performance.

Faster time to value: Public cloud offers a fast onramp to Generative AI to test and experiment. Private cloud offers many options to tailor models to your use cases, with more control.

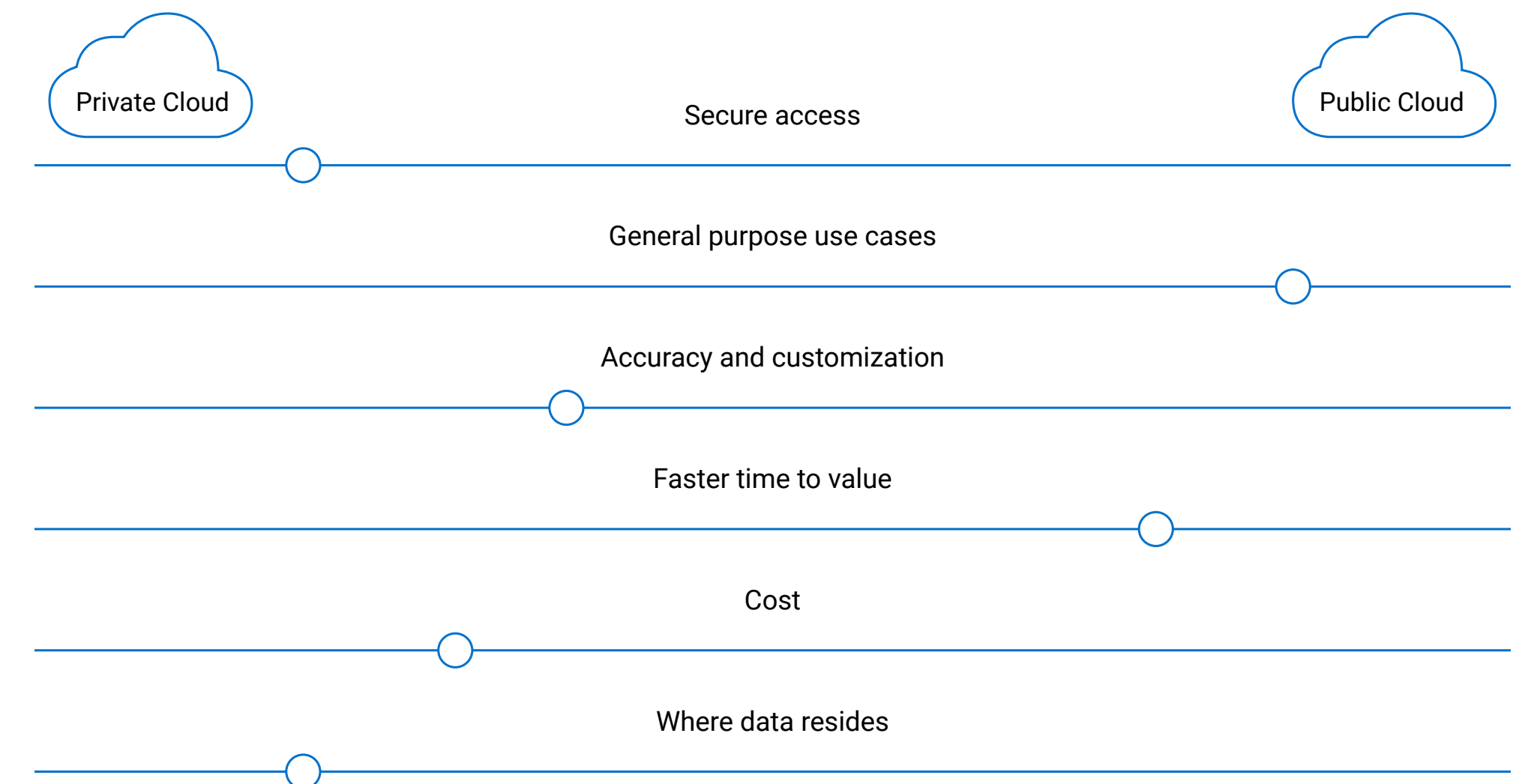
Cost: Private cloud offers more transparency and the ability to right-size models, so you have more control over costs.

Where data resides: AI works best when it's paired with your data wherever it resides, whether it's at the edge, in the cloud, or on premises.

5 reasons to bring AI to your data

The importance of data in the AI era cannot be overstated, and unlocking Generative AI's full potential depends on your ability to leverage it alongside your own proprietary data. Here are five reasons this is beneficial:⁵

1. Control security and access to your data.
2. Create more guardrails and reduce reputational risk.
3. Capitalize on your own real-time data.
4. Develop cost efficiencies.
5. Become more energy efficient.



Generative AI and data management

What's your data management strategy?

Most organizations are taking a two-pronged approach to their Generative AI strategy. They're experimenting with tactical deployments to learn and avoid falling behind, while also developing a long-term strategy that's flexible enough to accommodate the innumerable use cases that will emerge over time. This approach requires a two-tiered data management strategy as well.

Long term: Organizations must have a proper data management strategy for Generative AI, addressing key concerns such as data governance, data quality, data integrity, and data security. This is not just a technological concern, but an organizational one, as many of these concerns will require close collaboration with business stakeholders.

Organizations will also need a repository such as a data lake or data lakehouse for their Generative AI data, one that provides the controls the business expects as well as the flexibility to combine data sets as needed for Generative AI use cases and associated models.

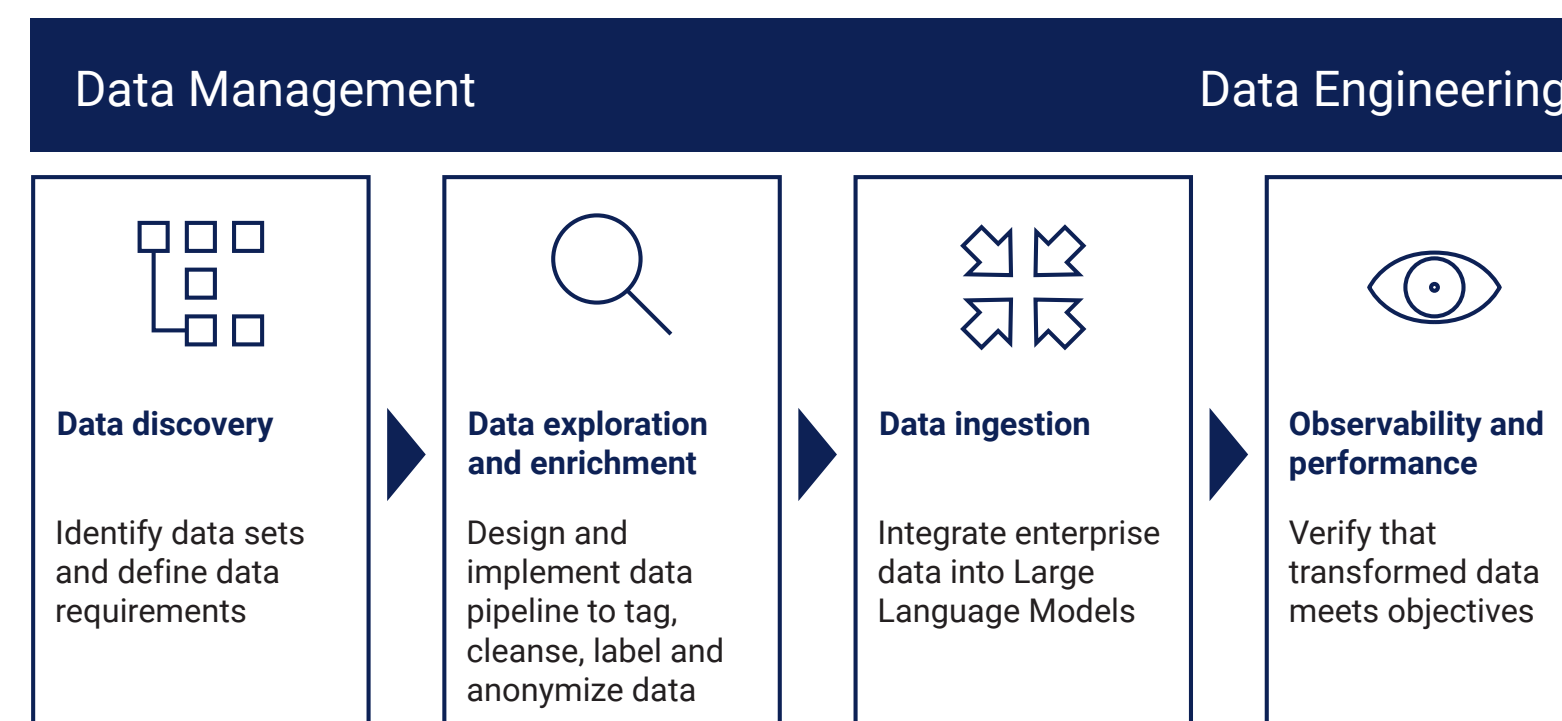
It's important to build these data management capabilities iteratively, enabling simultaneous development of both data management and Generative AI skills, and progressively expanding the capabilities of the foundational Generative AI data repository.

Dell Technologies data preparation services

Short term: Organizations are eager to get Generative AI use cases up and running to start capturing the value that Generative AI models can bring to their data. This fast-track approach requires data preparation, which includes:

- Anonymizing data to protect against exposing personal information
- Labeling the data
- Normalizing it across data sources.

It also requires building data pipelines to integrate the data into a model. Dell has dedicated Data Preparation Services for Generative AI to guide organizations through each of the crucial steps and to accelerate your time to value for high-priority use cases.



The way forward to maximize the value of these LLMs is to combine and train on your own data, based on the use case. This all starts with a strong data management strategy to ensure the quality of your data, and then ladders down into planning out the complex process of customizing the model—as most organizations' use cases require combining domain-specific information from their vertical into a pre-existing large language model.

—Jonathan Seckler,
Dell Technologies

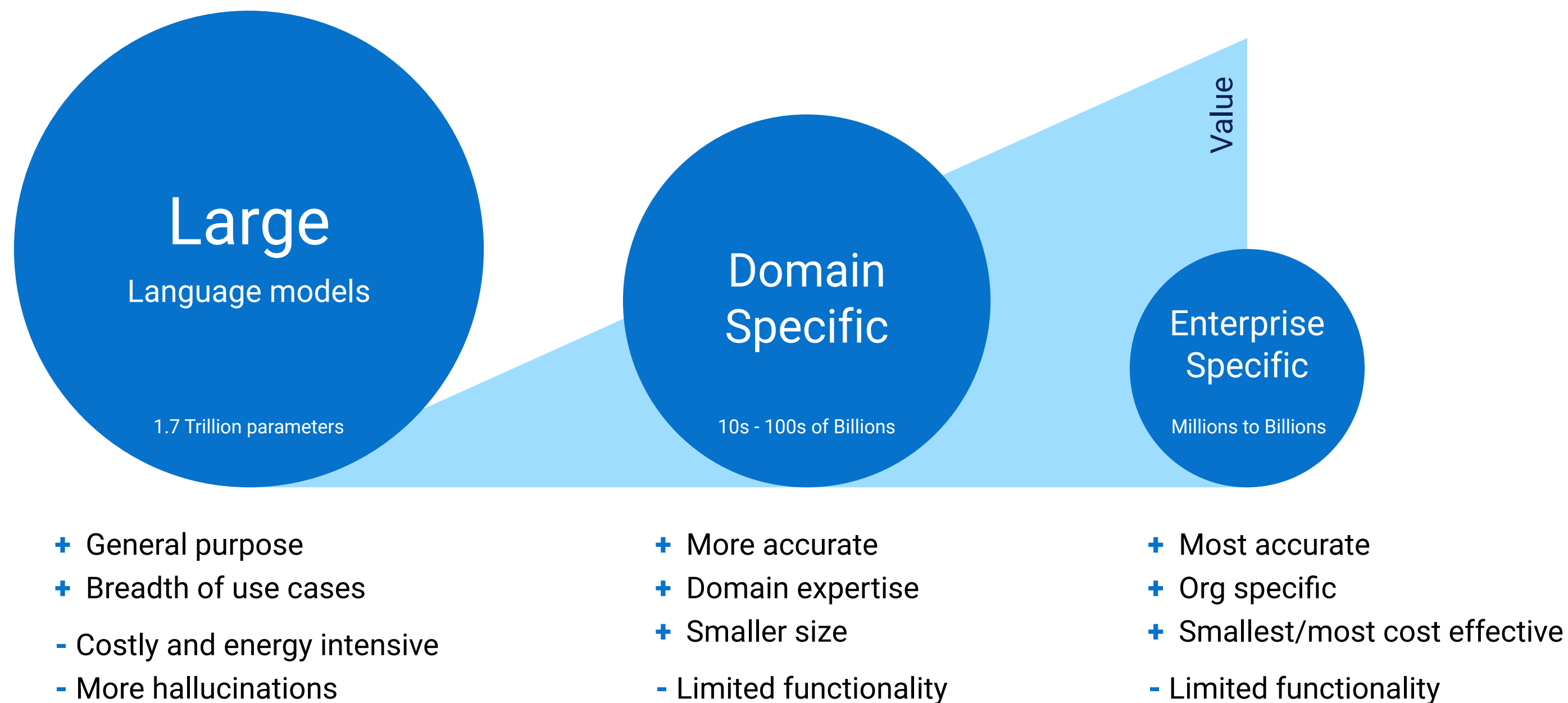
Right-sizing your generative AI models

Generative AI solutions come in a spectrum of sizes

Different types of Generative AI have corresponding entry points and investments that are necessary to ensure success. LLMs trained on vast amounts of text are like encyclopedias. They are terrific for general use but not for specific questions regarding your organizational data.

Your data is extremely powerful and unique. The more you leverage it, the more you can lean into domain and enterprise-specific use cases with focused industry value involving particular tasks or functions that you alone have data on.

Your data improves the efficiency and value of Generative AI



Align your model to your use case for enhanced efficiency

Explicitly defining your use cases makes it easier to right size your model and avoid wasting resources on a model that is larger than you require. Smaller models are more cost-conscious and energy-efficient. Right-sizing your LLM for use on a workstation, a laptop, or at the edge is also helpful.

Forrester predicts that Generative AI will further reduce energy use in several ways⁴:

Measurement and reporting: AI can introduce automation to generate timely recommendations toward net-zero carbon pathways.

Governance and visualization: Generative AI is an ideal use case for a single source of truth, which requires fewer resources to manage.

Research: AI can help accelerate revolutionary approaches to creating less energy-intensive technologies and processes.

AI deployment models: Assessing cost and value tradeoffs

The first three types of deployment models below are what 90% of organizations are implementing now. The AI model you choose will depend on your organization's level of data science readiness, deployment patterns, and the implications of each. Model augmentation is a great place to start, and most organizations move from that phase to fine-tuning models.

Simple inferencing with a pre-trained model

Known as "prompt engineering," which asks a pre-trained model a question and receives a result. Using ChatGPT is an example of simple inferencing.

Model augmentation

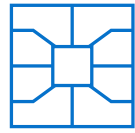
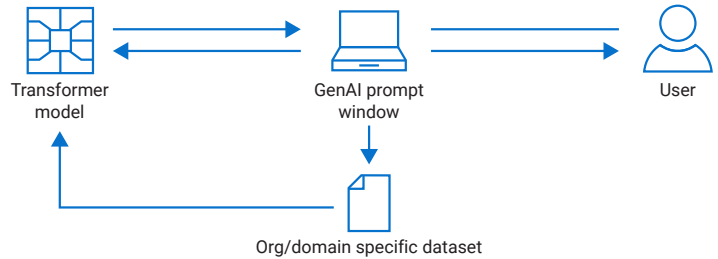
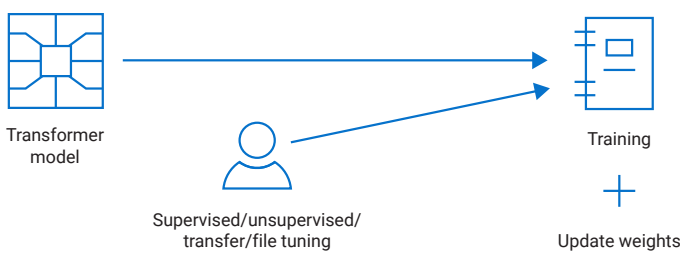
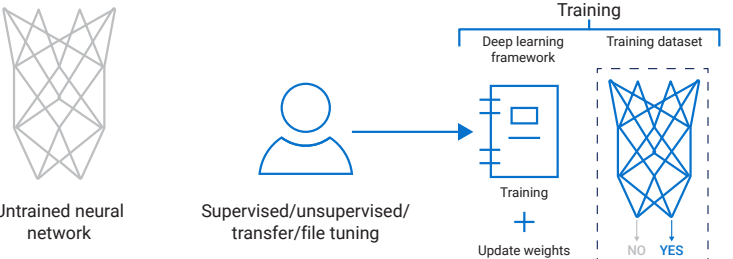
















Inferencing plus access to your data, for which RAG is a common use case. It's an easy onramp to making a Generative AI model more intelligent based on your data.






Fine-tuning models

Involves changing the weighting of a model and informing it with your data. Leveraged as organizations start to scale their models, it delivers better results but also requires more effort to establish.

Model training

Includes building a very specific model and training it with a data set. This typically requires the most work and resources and is often reserved for solving the most complex problems.

	Pre-trained model	Model augmentation (RAG, PAL)	Fine-tuning models	Model training
Model definition	Initially trained or pre-trained model  Transformer model			
Effort				
Cost				
Value & differentiation				
Data integration				
Infrastructure	Client – server	Client – server	GPU optimized	Large GPU deployment
Skills	IT Ops	Developer	Data scientist(s)	Data scientist(s)

-  Minimal amount
-  Low amount
-  Medium amount
-  High amount
-  Very high amount

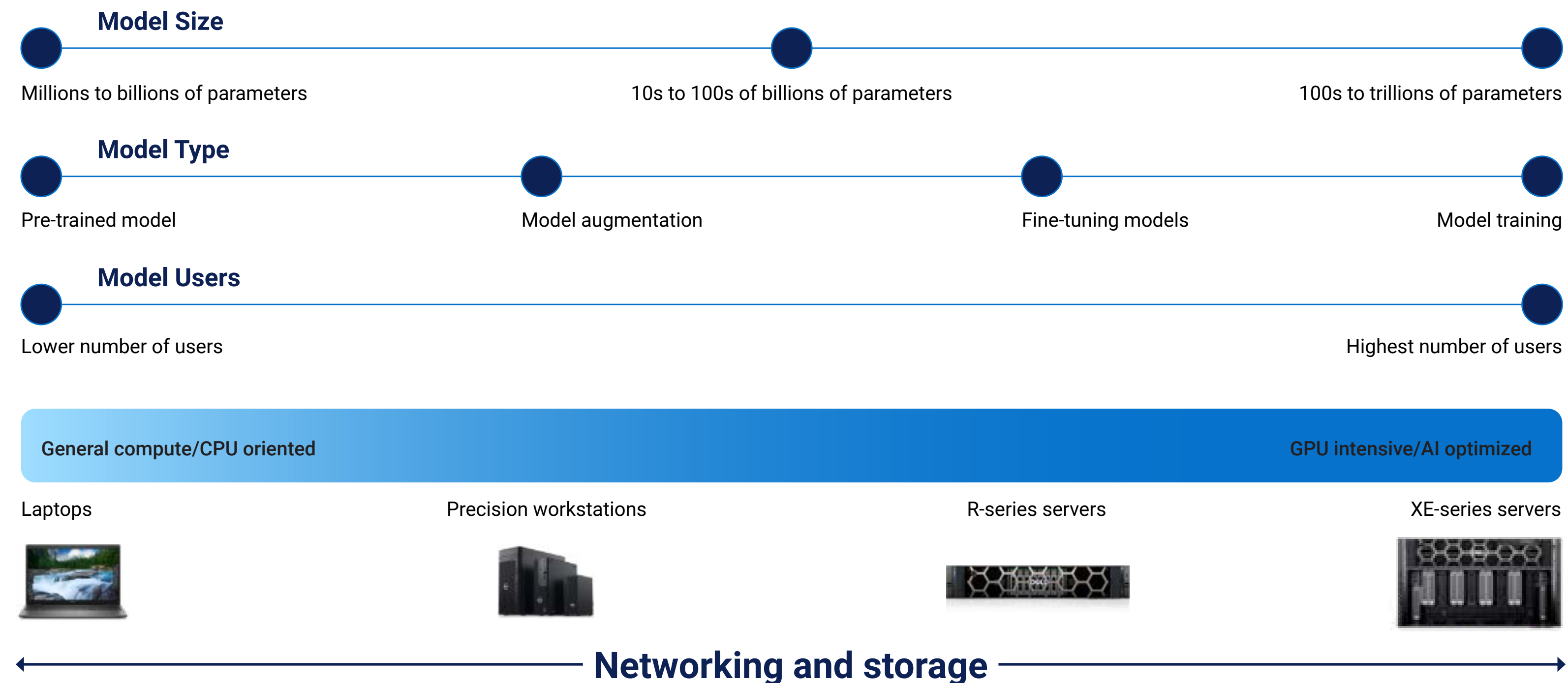
Having the right infrastructure for your models

Hardware needs increase along with the computational load

Simple inferencing use cases require a minimum investment of time and existing hardware. Model augmentation requires more compute-intensive infrastructure while fine-tuning models and model training call for increasingly higher compute power to run workloads. Your most compute-

intensive use cases will likely require a data scientist and high-performance AI-optimized servers.

As your Generative AI capabilities and use cases mature, you may need to right-size your infrastructure to optimize outcomes.



Bringing AI to your teams

Make Generative AI a compelling user experience

From a customer and business perspective, the most important step in Generative AI is the “last mile” of the process. You’ve created your Generative AI instances and right-sized your models, and now you need to roll out Generative AI broadly across your workforce, which will likely include technical and non-technical users. It’s time to help your workers understand the opportunities Generative AI brings to them.

Build a framework to help teams identify use cases

Help your teams with tools and resources such as templates they can leverage to build creative and intelligent solutions for their own practices, and then share these organization-wide. This helps Generative AI become a welcome and transformative process in your organization.

Prompt engineering and prompt libraries

Prompt engineering might become a fascinating new career option for some of your astute employees. Selecting the right words, phrases, or symbols that guide the AI model to generate relevant, high-quality texts is a critical skill for developing impactful AI-powered applications.

Training your workforce in prompt engineering practices can enhance productivity throughout your organization.



The last mile

Create organization-wide focus for Generative AI

Create and implement training to grow excitement about the possibilities of Generative AI across organization, and identify Generative AI heroes who can be trainers.

Centralize best practices

Develop a single source of truth for Generative AI knowledge, use cases, and best practices, helping your teams understand the governance and security measures so mistakes can be avoided.

Roll out enablement across organizations

Distribute Generative AI training and best practices across the organization to establish familiarity and trust in the technology. Group teams together based on skillsets, roles, and technical aptitude. Make sure you have a “Generative AI Captain” within each group to lead development.

Helping users get the most from Generative AI

Generative AI tools work a lot like chatbots

Users input questions, known as prompts, and the tool outputs answers in the form of text, code, images, etc. To get the best results, prompts should be aligned with business use cases and be specific as possible about the desired outcome. Here are three rules of thumb when helping users build effective prompts:

Tell the AI model what role it should play

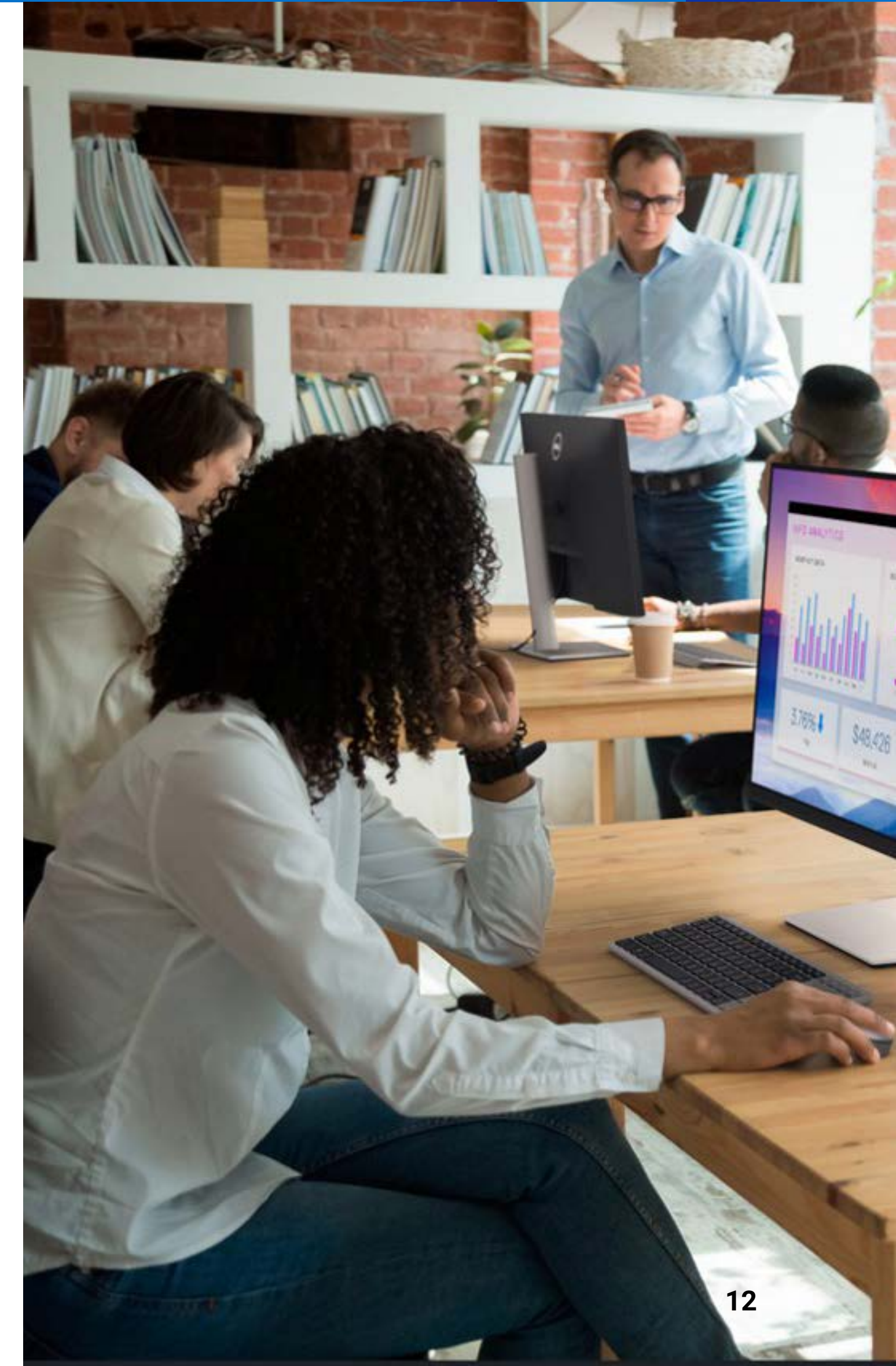
This helps the AI model better understand what results you might want. Examples include: expert marketer, copywriter, or social media expert.

Explain what you need

Clearly state what deliverable you’re looking for. For example, you might say article, blog post, social post, or web page copy. Or you might simply say something like “list”, “summary” or “bullet points”.

Provide specific, relevant instructions

This gives the AI model constraints that help it focus on the right output. Some examples might include: intended audience, style or tone to emulate, keywords to use or avoid, and desired length.



New beginnings

Generative AI: Ushering in a new era of productivity and innovation

The world currently stands on the precipice of epochal transformation. Generative AI is emerging as the beacon of this transformative journey, illuminating pathways to unprecedented productivity and innovation within the ever-evolving tableau of the future of work.

This breakthrough technology is not just an evolutionary step, but a leap heralding a new era where the boundaries of what we once deemed possible incessantly expand.

Break the barriers of productivity

Generative AI is extending the boundaries of human productivity; this technology operates around-the-clock without breaks or downtime. It can effortlessly scale up or down in response to workload fluctuations, offering businesses an unparalleled level of flexibility and efficiency.⁵ This constant, unwavering productivity is set to redefine efficiency across industries. Plan your strategy today to elevate business efficiencies.

Bridge the skill gaps of the future

Envision a future where Generative AI becomes the great equalizer in the workforce. The digital future continues evolving and the chasm between existing skills and industries' demands

widens. Traditional training and development initiatives have struggled with bridging this gap. Generative AI levels the playing field with personalized learning experiences tailored to an individual's unique learning style and pace.

Software vendors are already leveraging Generative AI in their productivity suites; this practice will continue to evolve and further innovation in the workplace. This technology emancipates employees from routine tasks, propelling them toward higher-value activities that foster intellectual development and innovation. This not only enriches the work experience but also equips the workforce with new skills and competencies, ensuring their relevance in the ever-evolving job market.

Unleash innovation faster

But perhaps the most exciting aspect of Generative AI lies in its capacity to catalyze innovation. It serves as an empowering platform for anyone with a vision and enables businesses of every size and structure to revolutionize an industry.

Generative design, a subset of Generative AI, empowers organizations to explore numerous design permutations swiftly, optimizing products for performance, aesthetics, and cost-effectiveness. This accelerates development cycles and enables companies to respond to market demands and trends, stay ahead of competition, and innovate at a rapid pace.



Getting started with Generative AI

The path to success is early wins

Retrieval-augmented generation (RAG) is an ideal early use case for many organizations that taps additional resources, such as your own data, to augment a model without the need to retrain it. RAG is used in LLM applications to retrieve relevant “knowledge base” style content, augment the user prompt with this domain-specific content, and then feed both the prompt and content into the LLM to generate a more complete, useful response.

You can run RAG use cases within your own firewall without the need for a data lake’s worth of data to get a proof-of-concept (POC) RAG use case up and running. This is a highly leveraged RAG use case that generates natural language summaries based on a user’s query and search/retrieval of corporate documents, which is a prime example of leveraging the value of your own data.

Most organizations have hundreds of use cases for which RAG can be leveraged. Here are some simple steps for setting up on-premises RAG use cases:

- Finding hardware for the trial
- Identify the use cases you want to solve so you can identify the right data
- Download open source software that turns unstructured data into structured data
- Create the Vector database that will house all the data you want the LLM to use
- Select an LLM and its parameter size; this impacts accuracy and hardware required.
- Now go have some fun prompting.

[Read our blog](#) to learn more about how to get started and how to access our source code for RAG.



Working with Dell Technologies, a renowned technology leader, will further accelerate our impact by combining our expertise in AI cloud infrastructure with Dell's global reach and resources. Together, we are poised to revolutionize the AI landscape and empower businesses worldwide to harness the true power of AI for a brighter, more sustainable future.

— Dave King,
Co-founder and CRO
Denvr Dataworks

Dell Technologies can help accelerate your journey

Learn more about how you can bring AI to your data and successfully create and implement Generative AI projects of all shapes and sizes. We've put together several resources to help you get started.

Adoption services and solutions

Dell Technologies can help you remove barriers and enable enterprise-wide Generative AI adoption through an end-to-end holistic approach from the desk side to the data center.

You can unlock a comprehensive Generative AI services portfolio to help your business and IT leaders map out strategies to expose the value from data silos and data lakes, close skillset gaps, and determine the right infrastructure for your Generative AI future.

We also provide a single source of customer support for every employee and use case. On a global scale and with global partners, Dell's Generative AI services and Generative AI solutions can help you focus on emerging opportunities and speed your time to tangible value with your Generative AI projects.

More Generative AI resources

- [Bring AI to Your Data with Dell Generative AI Solutions](#)
- [AI Technologies Powered by Dell Precision Workstations](#)



Get started on your Generative AI strategy in a fee-waived 1/2 day facilitated strategy conversation to help identify your key opportunities, challenges, and priorities.

[Contact us to set up your workshop >](#)