

## How to Pilot Generative AI

Published 10 July 2023 - ID G00797246 - 25 min read

By Analyst(s): Leinar Ramos, Anthony Mullen, Rajesh Kandaswamy, Radu Miclaus, Erick Brethenoux, Avivah Litan, Haritha Khandabattu

Initiatives: [Artificial Intelligence](#); [Evolve Technology and Process Capabilities to Support D&A](#); [Generative AI Resource Center](#)

When organizations begin experimenting with generative AI, they often neglect its transformative nature and the operationalization challenges it entails. IT leaders responsible for AI can follow the five steps in this research to pilot generative AI and start delivering business value.

### More on This Topic

This is part of an in-depth collection of research. See the collection:

- [Research Roundup for Generative AI](#)

## Overview

### Key Findings

- The most successful pilots focus on demonstrating business potential, not on technical feasibility. Organizations tend to run technical pilots that simply demonstrate that it is possible to build something with generative AI, leading to only incremental improvements and ignoring the transformative potential of this technology.
- IT leaders struggle to identify and prioritize impactful generative AI use cases due to the broad and emerging nature of the technology.
- Mature AI organizations involve business partners and software engineers as key members of their AI projects and pilot teams.
- Generative AI allows for faster development cycles than traditional AI projects. For this reason, a generative AI pilot requires a lean cycle of innovation — short experiments to test how the technology could add strategic value — while mitigating the potential risks that come with generative AI.
- Success in generative AI pilots requires rapid testing, refinement and, often, the elimination of use cases that do not have the anticipated effect on business value.

### Recommendations

As an IT leader focused on leveraging generative AI to create business value, you should:

- Run a workshop to generate use-case ideas with the business, focusing on the disruptive potential of generative AI and the way in which it can enable strategic objectives.
- Prioritize the use cases for your pilot against their potential business value and feasibility. Focus on no more than a few use cases for your generative AI pilot.
- Assemble a small but diverse team, including business partners, software developers and AI experts. Dedicate this fusion team for the duration of the pilot.
- Create a minimum viable product to validate each use case. Identify the target business key performance indicator (KPI) improvement hypothesis, and define the deployment approaches and risk mitigations required to quickly test this hypothesis.

- Deliver the minimum functionality required to test the use cases, and refine your assumptions on the cost and value of scaling them. Decide whether to stop, refine or scale each use case. Build upon initial successes to expand the generative AI pilot.

## Introduction

Organizations are planning to implement generative AI. In a recent Gartner webinar, 45% organizations have seen an increase in AI investment since ChatGPT was released, and 68% of executives believe that generative AI benefits outweigh its risks. However, despite this excitement, organizations struggle to get started: Only 19% of organizations are currently running a pilot or producing generative AI (see [Executive Pulse: AI Investment Gets a Boost From ChatGPT Hype](#)).

*Generative AI can be overwhelming.* It opens up a broad set of opportunities: Organizations can use it to generate many different types of artifacts, including text, code, images, video, music, speech and designs (e.g., 3D, parts and buildings).

A generative AI pilot might cover a diverse set of use cases, for instance:

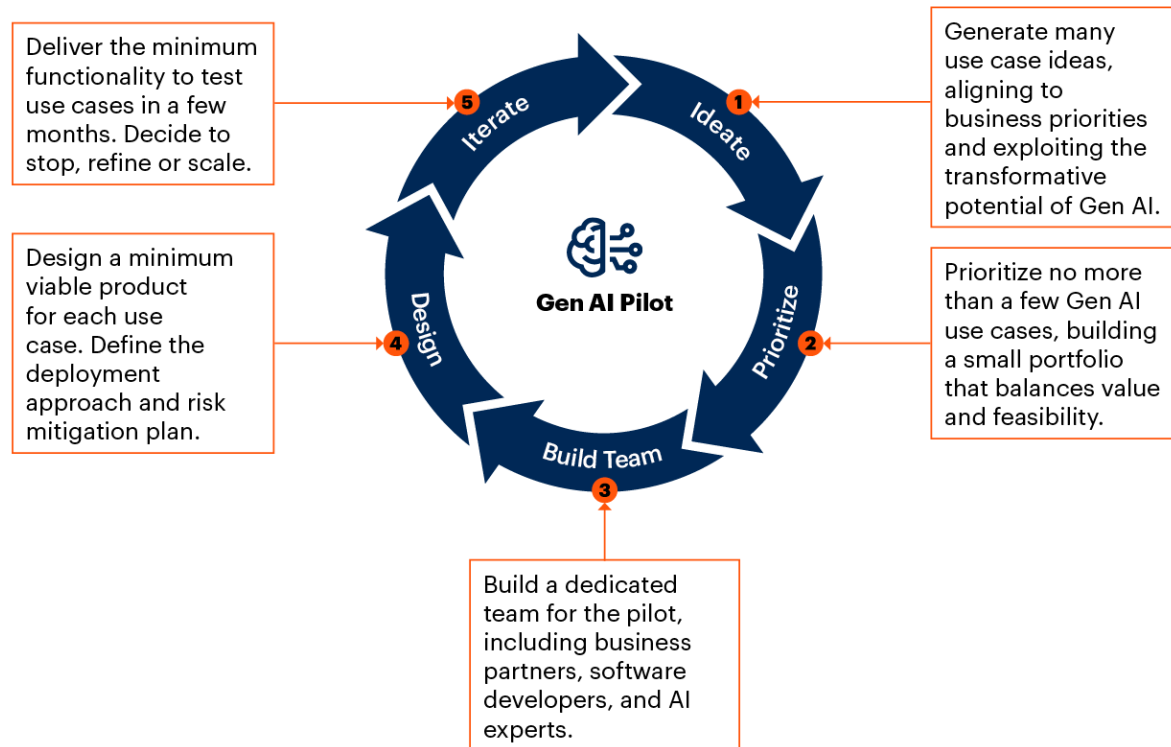
- Applying large language models (LLMs) to create drafts for agents to reply to customer inquiries
- Using a code-generation application to improve developer productivity
- Leveraging image generation to create marketing materials

Moreover, there are different approaches to implementing this wide range of use cases, from buying an external application and customizing foundation models all the way to building your own AI models from scratch.

Given this complexity, what steps can IT leaders take to initiate and run a generative AI pilot?

IT leaders need a systematic way to navigate the wide space of opportunities. They must find and prioritize a few of the most promising use cases for their organization and run a pilot to quickly deliver and validate these ideas in practice. To achieve this, IT leaders can follow the five steps described in this research (see Figure 1).

Figure 1: Generative AI Pilot Cycle

**Generative AI Pilot Cycle**

Source: Gartner  
797246\_C

Gartner

**Analysis****Step 1: Identify Transformative Use Cases With the Business**

When a new, transformative technology like generative AI (GenAI) is first introduced, the temptation is to run a technical proof of concept that simply demonstrates that it is possible to build something with it. This results in only incremental improvements and leaves IT leaders with a false sense of comfort because they have experimented with the technology. *This is a mistake.*

**Your generative AI pilot should not be about just proving that the technology works, but about learning how generative AI fits into the future of your company.**




Understand How Generative AI Can Transform the Organization

The goal of your pilot should be to learn how generative AI can drive strategic value, so it is key to focus on its transformation potential. There are three generative AI disruption patterns that go beyond what was previously possible with other AI techniques (see Figure 2):

- **Content consumption:** Customers and employees will increasingly expect their content consumption to be enhanced with generative AI via chat and similar interfaces. Enterprises will need to enable this new user experience through UX design, knowledge management and search.
- **Content generation:** Employees will use AI-powered assistants to complete most of their tasks. There will be a disruptive shift across multiple areas of content generation – whether it is generating text for internal communications and external messaging or generating visual creative content across a wide spectrum of modalities (image, video and sound).
- **Technology creation:** Generative AI will accelerate technology creation, not only improving the productivity of developers, but also democratizing technology creation further by enabling business technologists to more easily develop applications.

Figure 2: Key Generative AI Disruptions

Key Generative AI Disruptions

	Current State	Generative AI
 Content Consumption	Specialized skills required to consume data and knowledge	Information accessed in natural language and presented in a compelling way.
 Content Generation	AI used for predictive analytics, automating tasks, classification and prediction	AI used for generating many artifacts (such as text, images, code, video, audio & data).
 Technology Creation	Concentrated in a few specialized resources	<ul style="list-style-type: none"><li>• Accelerated technology creation.</li><li>• Sophisticated technology can be built by nontechnologists.</li></ul>

Source: Gartner  
797246\_C

These are broad categories of generative AI opportunities, but they need to be translated into specific use cases for your organization. For this, it is crucial to involve the business from the very start of your generative AI journey: *Without business involvement, your pilot will become an IT demo.*

**Capitalize on the broad excitement around generative AI. Involve business partners from the very start of your pilot.**

## Run Ideation Workshops With the Business

To involve the business, first identify and partner with a senior executive sponsor for the pilot. The ideal candidate would have both an interest in the topic and enough influence across the organization to get the pilot going and address potential barriers.

After gaining executive support, organize an ideation workshop to discuss how generative AI can enable your strategic business goals. This should include a mix of business partners, AI and IT experts. The objective is to generate ideas for use cases and gain buy-in for an initial pilot. This session should be short and targeted.

Start the workshop by creating a shared understanding of the technology, presenting a balanced view of opportunities and risks (see [Board Brief on Generative AI](#)). Introduce the three disruption patterns from Figure 1 to communicate how generative AI enables new-in-kind use cases. Then, run an exercise to generate potential use cases.

To maintain a focus on the strategic potential of the technology, you can follow the format in Table 1, encouraging participants to think about the strategic goals and how generative AI and specific use cases could contribute. IT leaders should push for the participants to think beyond incremental improvements to their processes and to imagine how generative AI could lead to transforming the business.

**Table 1: Format and Example of Ideation Exercise**

(Enlarged table in Appendix)

<i>Business Goal</i> ↓	<i>How Generative AI Can Transform (Mapped to Disruption Patterns)</i>	↓ <i>Sample Use Cases</i> ↓
Improved Customer Satisfaction	Content consumption: Customers will increasingly expect chat and similar generative AI interfaces to interact. They will expect personalized and immediate responses.	<ul style="list-style-type: none"> <li>■ Drafting customer service responses using text generation</li> <li>■ Customer-facing chatbot</li> <li>■ Sentiment analysis based on previous customer interactions</li> </ul>
Increased Employee Productivity	<p>Technology creation: Generative AI will accelerate technology creation, not only improving developer productivity, but also democratizing technology creation.</p> <p>Content generation: Employees will be supported with AI-powered assistants for many of their tasks (coding, design, customer service interactions, etc.).</p>	<ul style="list-style-type: none"> <li>■ Code generation to improve developer productivity</li> <li>■ Business intelligence (BI) and analytics co-pilot to make it easier to generate and consume BI</li> <li>■ Employee chatbot linked to knowledge repository</li> <li>■ Support for designers (image-generation assisted)</li> <li>■ Search and summarize legal documents</li> </ul>
Top-Line Revenue Growth	Content generation: Create personalized campaigns.	<ul style="list-style-type: none"> <li>■ Image generation for marketing materials</li> <li>■ Drafting product descriptions</li> <li>■ Creating personalized recommendations</li> </ul>

Source: Gartner (July 2023)

To expand idea generation:

- Run a campaign to collect ideas broadly from across the organization. Keep this simple, with a clear form to collect ideas and incentives for people to contribute.
- Identify what other organizations are doing in terms of generative AI use cases (see [Tool: Enterprise Use Cases for ChatGPT](#)).

The focus in this first stage is to collect as many ideas as possible without prioritizing them: *Use case prioritization should be kept separate from idea generation.*

## Step 2: Prioritize the Use Cases for Your Pilot

If done well, the idea-generation step should result in dozens of potential generative AI use cases. However, Gartner does not recommend running many use cases at once when starting to experiment with this technology.

The key to pilot success is prioritization: *Select no more than a few generative AI use cases to start with for the short duration of the pilot.*

Prioritize your use-case ideas by scoring them according to their business value and feasibility:

- **Business value:** How much does this particular use case contribute to each of the strategic objectives of your organization, such as those defined in Table 1 (e.g., improved customer satisfaction, increased employee productivity and top-line revenue growth)?
- **Feasibility:** What is the risk profile of this particular use of generative AI (see Table 2 for generative AI risks)? Is this use case currently feasible with the technology available? <sup>1</sup> Are there any turnkey solutions that will make it easier to implement (see Deployment Approaches section in Step 4)? How ready is your organization to adopt it?

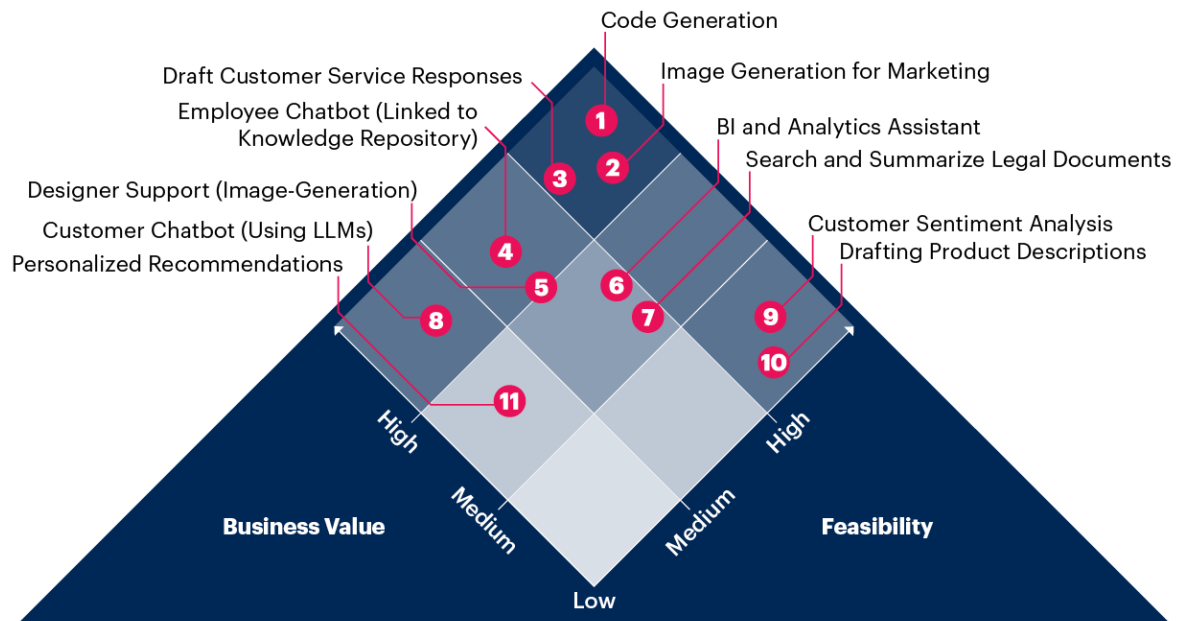
**Start with generative AI use cases that do not require sharing confidential information or intellectual property, as well as those where output accuracy does not need to be perfect.**

You can follow Gartner's AI prism methodology, which enables you to define criteria for business value and feasibility, and then score AI use cases (see [Toolkit: Discover and Prioritize Your Best AI Use Cases With a Gartner Prism](#)). Figure 3 shows an *illustration* of what a prioritization might look like, continuing on the use-case examples given in Table 1. (Note: This example does not represent Gartner's view of the value and feasibility of these use cases.)



Figure 3: Illustrative Example of Generative AI Prioritization

### Illustrative Example of Generative AI Prioritization



Source: Gartner

BI = business intelligence; LLM = large language model

797246\_C

Gartner

In the example in Figure 3, the organization might decide to prioritize the three use cases at the top: code generation, image generation for marketing, and the drafting of customer service responses via text generation.

Think about these initial use cases as a *portfolio*. Balance risk and reward, investing in some use cases that are more feasible, and others that are more valuable. Diversify these investments by selecting use cases that contribute to different business objectives.

**Take a small bite of a big business problem. Your pilot use case portfolio should deliver business value directly, but also demonstrate the strategic potential that generative AI could have in your organization.**

### Step 3: Assemble a Small but Diverse Team

The pilot team should include everyone required to run the full pilot. Most teams should include at least three types of team members (see Figure 4):

- **Business experts:** Involve experts from the business units that will be impacted by the selected use cases. They will help align the pilot with their strategic goals and identify any change management and process reengineering required to successfully run the pilot.
- **Software engineers:** Provide user-interface, front-end application and scalability support. The 2022 Gartner AI Use Case ROI Survey shows that organizations in which software engineers are involved in the stage of developing AI use cases are much more likely to reach mature levels of AI implementation. <sup>3</sup>
- **Data scientists and AI experts:** These contributors manage the AI components of the projects. You are unlikely to find generative AI experts in your organization because the field is relatively new, but look for staff members with deep learning skills, ideally with experience on transfer learning techniques and working with foundation models.

The degree of involvement of these roles may vary depending on the specific use cases that have been selected. However, this fusion team, combining IT and business experts, will be in a much better position to succeed than technical teams that consult only the business for their requirements (see [Fusion Teams: A Proven Model for Digital Delivery](#)).

Dedicate the people in this team for the full duration of the pilot (e.g., two or three months). Use your pilot as a way to develop the new skills required for generative AI (see [Quick Answer: How Will Prompt Engineering Impact the Work of Data Scientists](#)).

Beyond the core pilot team, you might need to involve other parts of the organization at different points of the pilot:

- A generative AI governance team, which should define policies and ensure they are implemented within the pilot
- Security and legal experts to assess risks and consider risk mitigations for the pilot (see the Determine Risk and Mitigation for Pilot Use Cases section in Step 4)
- Technical experts for testing and quality assurance for the pilot

- Enterprise architects to set and enforce standards in the deployment phase of pilot use cases

## Step 4: Design and Plan the Pilot

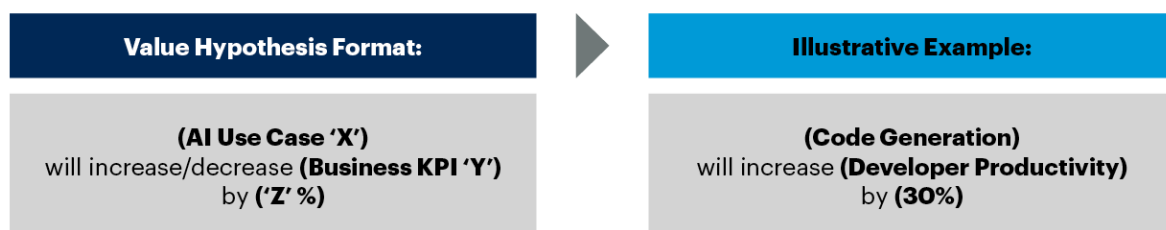
### Determine Pilot Objectives and KPIs

The pilot team must quickly go through a design and planning sprint that lasts only *one or two weeks*. The objective of this phase is *not* to design a fully-fledged application, but to plan for a minimum viable product (MVP) — that is, a product that requires the least amount of effort to validate a hypothesis of the value that it could create for customers or employees. Additional features will come later.

The AI team must define a value hypothesis, which is an assumption about the improvement that the AI use case will have on a specific business KPI. You can use the simple format in Figure 4 to document the expected impact on the KPI (see Step 2 in [Capture AI Value With These 5 Benefit Realization Best Practices](#) for more details).

**Figure 4: Value Hypothesis Format and Example**

### Value Hypothesis Format and Example



Source: Gartner  
797246\_C

**Gartner.**

This business KPI must be identified at the start of the project, without exception — this is key to success; the 2022 Gartner AI Use Case ROI Survey found that mature AI organizations are more likely to define their business metrics at the ideation phase of every AI use case than low-maturity ones. <sup>3</sup>

The pilot will be able to not only demonstrate the value hypothesis, but also validate assumptions about the costs for scaling each use case (see [Assess the Value and Cost of Generative AI With New Investment Criteria](#)).

### Determine Risks and Mitigations for Pilot Use Cases

Generative AI unlocks many use cases, but it also brings additional risks:

- **Output and input risks:** Generative AI's outputs can be unreliable given factual errors and hallucinations. Outputs can also be biased on and potentially include copyrighted material or other unwanted, illegitimate or illegal information. Inputs into generative AI models can result in data compromise risks.
- **Privacy and data protection risks:** Many generative AI implementations will depend on access to third-party foundation models and applications. The use of private, proprietary, sensitive or confidential information as inputs into these generative AI models entails risks of data leakage and potential violations of existing regulations. Organizations must monitor and enforce privacy, confidentiality and governance of their data in the environments where models are hosted.
- **AI application security risks:** AI introduces security threats that are not addressed by conventional security controls. These include prompt injection attacks and hacker access to model states and parameters. Further, attackers themselves can use generative AI for hacking techniques that enterprises are not prepared for, and this is independent of the new AI application attack vector.

Given this, the pilot team must set the risk mitigations and controls for their use cases. See Table 2 for a format for risk assessment and mitigation plan for each generative AI use case in the pilot scope.

**Table 2: Risk Assessment and Mitigation**

(Enlarged table in Appendix)

Risk Category	Risk	Risk Mitigations
Output and input risks	Unacceptable use that compromises enterprise decision making and confidentiality	Define policies for acceptable use and establish a system and process to record user requests to use GenAI applications, their intended use and data required, approvals by various constituents, and periodic user attestations relative to application request and usage. Use third-party or homegrown input content filtering for information submitted to hosted LLM environments, to screen inputs against enterprise policies for acceptable use.
Output and input risks	Confidential data compromise	Use legacy security controls based on data classification, including data protection, role-based access management and layered security to protect data leaving the enterprise network. Require users to access LLM applications through enterprise portals where access can be controlled and monitored with legacy security controls and systems.
Output and input risks	Inaccurate, hallucinations or illegal outputs	Treat LLM outputs as first drafts. Ensure thorough manual output review, especially in sensitive and higher-risk applications. Use third-party or homegrown output content filtering for information received from hosted LLM environments to screen outputs against enterprise policies for acceptable use and legal compliance (e.g., to prevent use of copyrighted content in LLM responses).
Privacy and data protection risks	Data leakage and potential compromise in hosted LLM vendor environment	Opt out of the hosting vendor's prompt history storage. Ensure the hosting vendor does not use enterprise data for training its models. Scrutinize hosting vendor licensing agreements to define the rules and liability governing responsibility for data protection in the LLM environment.
Privacy and data protection risks	Failure to meet regulatory compliance (e.g., for GDPR, EU AI Act)	Determine needs to comply with existing and new regulations in your jurisdictions, and ensure you can meet them with the hosted LLM solution chosen — for example, required privacy impact assessments and data residency/sovereignty requirements. Implement requirements accordingly — for example, ensure data locations are compliant and personally identifiable information (PII) data can be excluded from hosted LLM interactions.
Cybersecurity	Prompt injection attacks	Use security service edge and other legacy security solutions to inspect prompts for injections upon submission to LLM environments and upon responses back into the enterprise network. Hold hosting LLM vendors to account for how they prevent indirect prompt injection attacks directly into their LLM environments (upon inputs or outputs) for which the enterprise has no control or visibility.
Cybersecurity	Hacker access to model states and parameters	Implement strong security around the local enterprise LLM environment, including access management, data protection, network and endpoint security.

Source: Gartner (July 2023)

## Decide on the Deployment Approach

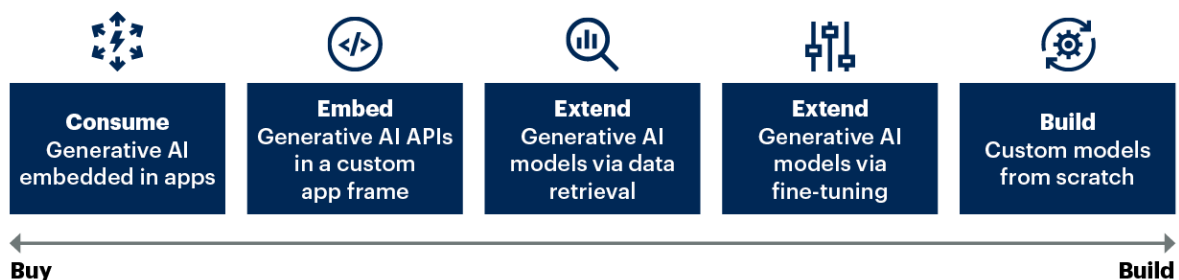
Considering both the pilot objectives and its risks, the pilot team must decide on the deployment approach for each use case. There is a variety of options to consume generative AI capabilities (see Figure 5):

1. **Consume generative AI embedded in applications:** Organizations can directly use commercial applications that have generative AI capabilities embedded within them. An example of this would be using a design software application that includes image-generation capabilities (e.g., Adobe Firefly).

2. **Embed generative AI APIs in a custom application:** Enterprises can build their own applications, integrating generative AI via foundation model APIs. Most closed-source generative AI models (GPT-3, GPT-4, PaLM 2, etc.) are available for deployment via cloud APIs.
3. **Extend generative AI models via data retrieval:** Foundation models can be augmented with data retrieved from additional document databases. This requires a retrieval/search step before using the generative models.
4. **Extend generative AI models via fine-tuning:** Fine-tuning further trains a large foundation model on a smaller dataset for a specific use case. For example, an insurance company could fine-tune a foundation model with its own policy documents to incorporate this knowledge into the model.
5. **Build custom foundation models from scratch:** Organizations could ultimately build their own foundation models from scratch, fully customizing them to their own data and business domains. This option is unfeasible for most organizations. For example, a financial institution might create a foundation model trained with financial data, which could then be used for many use cases.

**Figure 5: Generative AI Deployment Approaches**

### Generative AI Deployment Approaches



Source: Gartner  
797246\_C

**Gartner.**

There are trade-offs to each of the five options (see [How to Choose an Approach for Deploying Generative AI](#) for a full discussion).<sup>2</sup> However, because the objective of the pilot is to quickly validate the value hypothesis, the approaches of consuming applications or embedding APIs are preferable over the options of extending via fine-tuning or building foundation models from scratch.

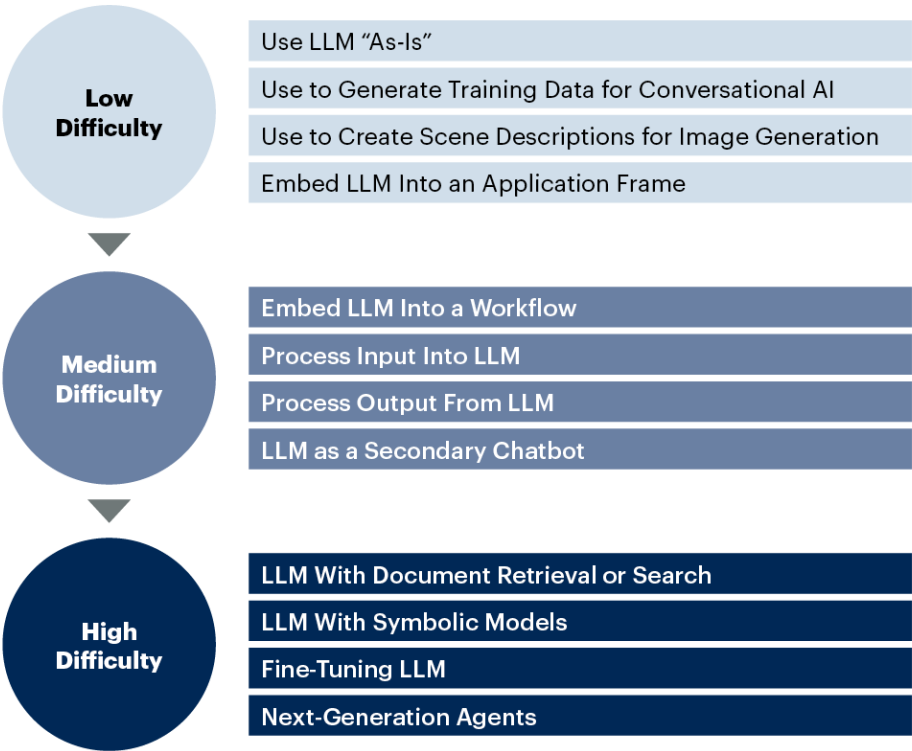
Don't start your generative AI pilot from scratch, use the foundation models and applications available externally.

Depending on the deployment approach, the team will be able to design the architecture of its initial use cases, selecting among different generative AI design patterns. See [AI Design Patterns for Large Language Models](#) to explore different ways to utilize generative AI foundation models.

Figure 6 illustrates different design patterns.

Figure 6: Design Patterns for Large Language Models (LLMs)

Design Patterns for Large Language Models (LLMs)



Source: Gartner  
797246\_C

The specific patterns chosen for each use case will depend on the objective defined, as well as the organizational readiness in terms of skills and data.

Each use case is therefore mapped to a deployment approach and design pattern (see Table 3).

Table 3: Illustrative Example of Use Cases Mapped to Design Patterns

<i>Use Case Example</i> ↓	<i>Deployment Approach</i>	<i>Design Pattern</i> ↓
Draft customer service responses	Extend generative AI models via data retrieval	LLM with document retrieval/search
Code generation to improve developer productivity	Buy a generative AI embedded application	Use application as-is
Image generation for marketing	Buy a generative AI embedded application	Use application as-is

Source: Gartner (July 2023)

Delivering the pilot use cases is likely to require selecting a vendor that offers the underlying foundation model capabilities — either via full-fledged applications or APIs. Contemplate whether your existing portfolio of applications and vendors could fulfill a role in the pilot because it might be faster to get started with an existing relationship. Regardless, consider at least the five criteria in Table 4 when selecting a vendor for your pilot.



**Table 4: Vendor Criteria for Generative AI Pilot**

(Enlarged table in Appendix)

Criteria ↓	Description ↓
Performance	Not all foundation models are created equal. It is important to assess and compare the performance of the underlying foundation models, ensuring the vendors have the type of model required for your use case (chat model, embedding model, etc.). This should include a consideration on the latency required for your use case.
Customizability	Depending on the selected design pattern, it is key to check if it is possible to customize the application and/or underlying foundation model. For instance: Is it possible to create customized workflows? Is it possible to inject your data via retrieval? Is it possible to fine-tune the underlying models? Could the vendor solution integrate with your own systems?
Privacy, Security and Data Retention	It is critical to establish legally binding data protection/privacy assurances in license agreements with vendors hosting the LLM models. Some vendors store the prompt data until a certain time for debugging purposes in the event of a failure or for investigating patterns of potential abuse and misuse. You must check if this data retention meets your own enterprise security policies.
Pricing	It is important to consider the different pricing models (e.g., subscription, API pay-as-you-go pricing and fine-tuning costs) as well as additional costs (cloud and AI infrastructure, implementation costs, etc.). The costs will vary depending on the use case, scale and requirements of the pilot.
Other Long-Term Considerations	There will be other considerations that are less important for the pilot, but much more important for a longer-term engagement, such as the maintenance and support offered.

Source: Gartner (July 2023)

The pilot is *not* the moment to make long-term vendor decisions — this is a much longer process. It is simply an opportunity to run a small proof of concept utilizing external capabilities. For more advice on vendor identification and selection, see [Toolkit: Vendor Identification and Selection Guide for AI and Data and Analytics \(D&A\) Service Providers](#).

## Step 5: Deliver and Iterate

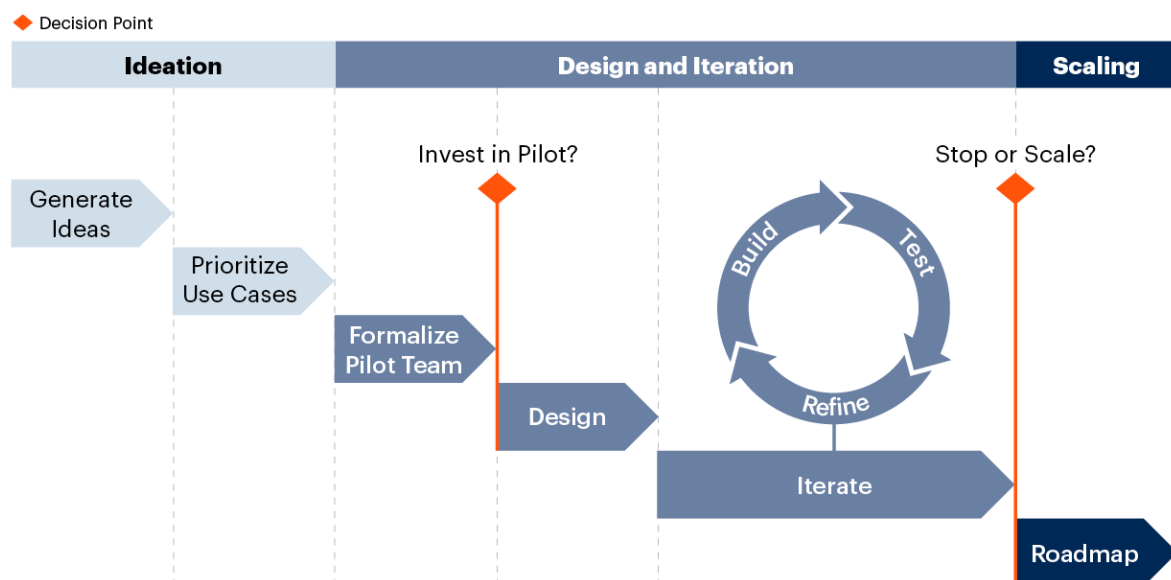
After the design and planning phase, the team must deliver on the pilot use cases. This is an iterative process in which teams will build, test and refine their implementation (see Figure 7). There should be, however, a time limit for this delivery phase: *Consider allowing no more than two months to refine and test each use case.*

This time constraint will force the team to focus on running the simplest pilot possible, testing the idea with the minimum functionality required and moving on to other ideas if the use case does not reach the intended performance in this period.

Generative AI allows for faster development cycles than traditional AI projects since teams can leverage pretrained foundation models. Use this speed to your advantage by quickly iterating and testing your pilot use cases.

Figure 7: Generative AI Pilot Phases and Decision Points

### Generative AI Pilot Phases and Decision Points



Source: Gartner  
797246\_C

Gartner

## Test and Refine Your Use Cases

Test the business KPI impact for each use case, as defined in the value hypothesis (see Step 4). The use case can be tested with a small number of users or customers to reduce risk exposure, but it's important to directly test it with the future users of the solution.

The initial use-case iteration will rarely meet the value hypothesis. However, the challenges encountered will be the crucial information to learn what needs to be improved.

Depending on the issues encountered, you might want to explore different options to improve performance (see [How to Improve the Performance of AI Projects](#)):

- **Modeling and architecture:** It could be that the performance of the GenAI model is simply not good enough for the intended use case. In this case, the team might explore ways to augment models with additional data or tools (see [AI Design Patterns for Large Language Models](#)) or consider alternative foundation models/applications that could bring higher performance.
- **Process engineering:** The issue might not be technical in nature, but instead driven by lack of user adoption. Efforts in change management and process engineering might be required in this case.
- **Data:** If you are using your own data (for example, when using a retrieval system), then you might consider improving data quality, increasing internal and external data collection, or improving data/feature engineering.

It could be the case, however, that despite the best efforts of the team after several iterations, it is not possible to achieve the intended business value for a given use case within the time limit. This is not a failure, but a signal that the real opportunities within generative AI might lie elsewhere. Discovering that a use case doesn't work at the pilot stage is a way to avoid the costs of building a long-term solution that won't be valuable for your organization.

**The goal of the pilot is to learn, not to deliver. Learn from early efforts and find a path forward.**

## Stop or Scale

By the end of this phase, the team will need to make a decision for each use case:

- **Stop:** There is no clear way to get the use case to the required performance. Document the lessons learned, including what worked and what did not. Go back to the ideation phase (Step 1) to consider other use cases not included in the original pilot.
- **Scale:** Having demonstrated the potential value of the use case, build a roadmap to scale, potentially expanding on the features from the pilot and deciding on the long-term architecture. The pilot phase should also uncover long-term feasibility and help to refine your assumptions on the total long-term-cost structure and investment required for delivering the solution at scale. To estimate the potential cost and return on investment of the initiative, see [Assess the Value and Cost of Generative AI With New Investment Criteria](#).

Build upon your initial success. Tell a story on the value obtained and lessons learned in your pilot, and expand into additional generative AI use cases by going back to the ideation phase (Step 1). Generative AI is a fast-moving field; new use cases will be unlocked, and the feasibility and value of use cases might also change as new techniques emerge and performance improves.

## Evidence

<sup>1</sup> The generative AI technology landscape is evolving quickly. Technical feasibility assessments should be made with a forward-looking approach.

<sup>2</sup> Open-source foundation models could be used in different deployment approaches. For example, they could be used as a starting point to fine-tune and create custom models.

<sup>3</sup> **2022 Gartner AI Use Case ROI Survey.** This survey sought to understand where organizations have been most successful in deploying AI use cases and figure out the most efficient indicators that they have established to measure those successes. The research was conducted online from 31 October through 19 December 2022 among 622 respondents from organizations in the U.S. (n = 304), France (n = 113), the U.K. (n = 106) and Germany (n = 99). Quotas were established for company sizes and for industries to ensure a good representation across the sample. Organizations were required to have developed AI to participate. Respondents were required to be in a manager role or above and have a high level of involvement with the measuring stage and at least one stage of the life cycle from ideating to testing AI use cases. Disclaimer: The results of this survey do not represent global findings or the market as a whole, but reflect the sentiments of the respondents and companies surveyed.

## Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

[AI Design Patterns for Large Language Models](#)

[AI Design Patterns for Knowledge Graphs and Generative AI](#)

[Assess the Value and Cost of Generative AI With New Investment Criteria](#)

[How to Optimize Enterprise Value From Data and Analytics](#)

[Tool: Enterprise Use Cases for ChatGPT](#)

---

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by [Gartner's Usage Policy](#). Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "[Guiding Principles on Independence and Objectivity](#)." Gartner research may not be used as input into or for the training or development of generative artificial intelligence, machine learning, algorithms, software, or related technologies.

Table 1: Format and Example of Ideation Exercise

<i>Business Goal</i> ↓	<i>How Generative AI Can Transform (Mapped to Disruption Patterns)</i>	↓ <i>Sample Use Cases</i> ↓
Improved Customer Satisfaction	<b>Content consumption:</b> Customers will increasingly expect chat and similar generative AI interfaces to interact. They will expect personalized and immediate responses.	<ul style="list-style-type: none"> <li>■ Drafting customer service responses using text generation</li> <li>■ Customer-facing chatbot</li> <li>■ Sentiment analysis based on previous customer interactions</li> </ul>
Increased Employee Productivity	<p><b>Technology creation:</b> Generative AI will accelerate technology creation, not only improving developer productivity, but also democratizing technology creation.</p> <p><b>Content generation:</b> Employees will be supported with AI-powered assistants for many of their tasks (coding, design, customer service interactions, etc.).</p>	<ul style="list-style-type: none"> <li>■ Code generation to improve developer productivity</li> <li>■ Business intelligence (BI) and analytics co-pilot to make it easier to generate and consume BI</li> <li>■ Employee chatbot linked to knowledge repository</li> <li>■ Support for designers (image-generation assisted)</li> <li>■ Search and summarize legal documents</li> </ul>

<i>Business Goal</i> ↓	<i>How Generative AI Can Transform (Mapped to Disruption Patterns)</i>	↓ <i>Sample Use Cases</i> ↓
Top-Line Revenue Growth	Content generation: Create personalized campaigns.	<ul style="list-style-type: none"><li>■ Image generation for marketing materials</li><li>■ Drafting product descriptions</li><li>■ Creating personalized recommendations</li></ul>

Source: Gartner (July 2023)

Table 2: Risk Assessment and Mitigation

<i>Risk Category</i> ↓	<i>Risk</i> ↓	<i>Risk Mitigations</i> ↓
Output and input risks	Unacceptable use that compromises enterprise decision making and confidentiality	<p>Define policies for acceptable use and establish a system and process to record user requests to use GenAI applications, their intended use and data required, approvals by various constituents, and periodic user attestations relative to application request and usage.</p> <p>Use third-party or homegrown input content filtering for information submitted to hosted LLM environments, to screen inputs against enterprise policies for acceptable use.</p>
Output and input risks	Confidential data compromise	<p>Use legacy security controls based on data classification, including data protection, role-based access management and layered security to protect data leaving the enterprise network.</p> <p>Require users to access LLM applications through enterprise portals where access can be controlled and monitored with legacy security controls and systems.</p>



<i>Risk Category</i> ↓	<i>Risk</i> ↓	<i>Risk Mitigations</i> ↓
Output and input risks	Inaccurate, hallucinations or illegal outputs	<p>Treat LLM outputs as first drafts. Ensure thorough manual output review, especially in sensitive and higher-risk applications.</p> <p>Use third-party or homegrown output content filtering for information received from hosted LLM environments to screen outputs against enterprise policies for acceptable use and legal compliance (e.g., to prevent use of copyrighted content in LLM responses).</p>
Privacy and data protection risks	Data leakage and potential compromise in hosted LLM vendor environment	<p>Opt out of the hosting vendor's prompt history storage. Ensure the hosting vendor does not use enterprise data for training its models.</p> <p>Scrutinize hosting vendor licensing agreements to define the rules and liability governing responsibility for data protection in the LLM environment.</p>

<i>Risk Category</i> ↓	<i>Risk</i> ↓	<i>Risk Mitigations</i> ↓
Privacy and data protection risks	Failure to meet regulatory compliance (e.g., for GDPR, EU AI Act)	Determine needs to comply with existing and new regulations in your jurisdictions, and ensure you can meet them with the hosted LLM solution chosen – for example, required privacy impact assessments and data residency/sovereignty requirements. Implement requirements accordingly – for example, ensure data locations are compliant and personally identifiable information (PII) data can be excluded from hosted LLM interactions.
Cybersecurity	Prompt injection attacks	Use security service edge and other legacy security solutions to inspect prompts for injections upon submission to LLM environments and upon responses back into the enterprise network. Hold hosting LLM vendors to account for how they prevent indirect prompt injection attacks directly into their LLM environments (upon inputs or outputs) for which the enterprise has no control or visibility.
Cybersecurity	Hacker access to model states and parameters	Implement strong security around the local enterprise LLM environment, including access management, data protection, network and endpoint security.

Source: Gartner (July 2023)

Table 3: Illustrative Example of Use Cases Mapped to Design Patterns

Use Case Example ↓	Deployment Approach ↓	Design Pattern ↓
Draft customer service responses	Extend generative AI models via data retrieval	LLM with document retrieval/search
Code generation to improve developer productivity	Buy a generative AI embedded application	Use application as-is
Image generation for marketing	Buy a generative AI embedded application	Use application as-is

Source: Gartner (July 2023)

Table 4: Vendor Criteria for Generative AI Pilot

Criteria ↓	Description ↓
Performance	Not all foundation models are created equal. It is important to assess and compare the performance of the underlying foundation models, ensuring the vendors have the type of model required for your use case (chat model, embedding model, etc.). This should include a consideration on the latency required for your use case.
Customizability	Depending on the selected design pattern, it is key to check if it is possible to customize the application and/or underlying foundation model. For instance: Is it possible to create customized workflows? Is it possible to inject your data via retrieval? Is it possible to fine-tune the underlying models? Could the vendor solution integrate with your own systems?
Privacy, Security and Data Retention	It is critical to establish legally binding data protection/privacy assurances in license agreements with vendors hosting the LLM models. Some vendors store the prompt data until a certain time for debugging purposes in the event of a failure or for investigating patterns of potential abuse and misuse. You must check if this data retention meets your own enterprise security policies.
Pricing	It is important to consider the different pricing models (e.g., subscription, API pay-as-you-go pricing and fine-tuning costs) as well as additional costs (cloud and AI infrastructure, implementation costs, etc.). The costs will vary depending on the use case, scale and requirements of the pilot.

Criteria ↓	Description ↓
Other Long-Term Considerations	There will be other considerations that are less important for the pilot, but much more important for a longer-term engagement, such as the maintenance and support offered.

Source: Gartner (July 2023)