The public sector guide to getting started with generative AI.

Table of contents

Google

Introduction

Generative AI marks one of the most significant technological shifts in history.

Generative AI marks one of the most significant technological shifts in history. Its impact on individual and organizational productivity can be significant, with the potential to rival the advent of the internet or the mobile device. Indeed, among organizations considering or using AI, 82% believe it will either significantly change or transform their industry.[1]

What makes generative AI different from other forms of AI that have come before is its ease of use in helping solve everyday problems in people's personal or professional lives. Anyone who knows how to ask a question of a search engine can use everyday language to interact with a generative AI chatbot or virtual agent — getting it to answer questions, create content, produce images, summarize documents, and much more.

Even better, a single generative AI platform can deliver solutions for multiple use cases, creating a network effect. As the number of users and applications increases, the model is exposed to more data and becomes increasingly accurate and useful — which in turn encourages more users.

Organizations that use generative AI to speed up, automate, scale, and improve operational processes stand to reap big benefits. According to McKinsey & Company, generative AI's impact on productivity could add between $2.6 trillion and $4.4 trillion USD annually to the global economy.[2]

Implications for public sector leaders

Government agencies and academic institutions hold a steadfast commitment to serving their mission and supporting the needs of their communities. While this commitment remains the same, the way that public sector leaders achieve the mission will continue to evolve – in large part due to AI. AI can help teams improve core services, solve problems, and deliver better value to meet the needs of today, tomorrow and beyond.

Google

This eBook is for agency leaders who want guidance on how to kickstart their organization's generative AI journey.

In the first chapter, you'll learn what generative AI is, what it is capable of, and the impact you can expect when applying it in an organizational context. The second chapter provides a step-by-step guide for getting started with generative AI, with recommended best practices from Google Cloud's AI experts and customers. In the third chapter, we dive into real-world examples of pioneers across the public sector who have adopted AI to work smarter, deliver value to constituents faster, and unlock new operational efficiencies. In the fourth and final chapter, learn how to choose the right solution to set your organization up for successful and responsible implementation. We include resources to use as needed on the last page.

With generative AI technology moving so fast, it can be overwhelming. We're here to help you chart the right path forward.
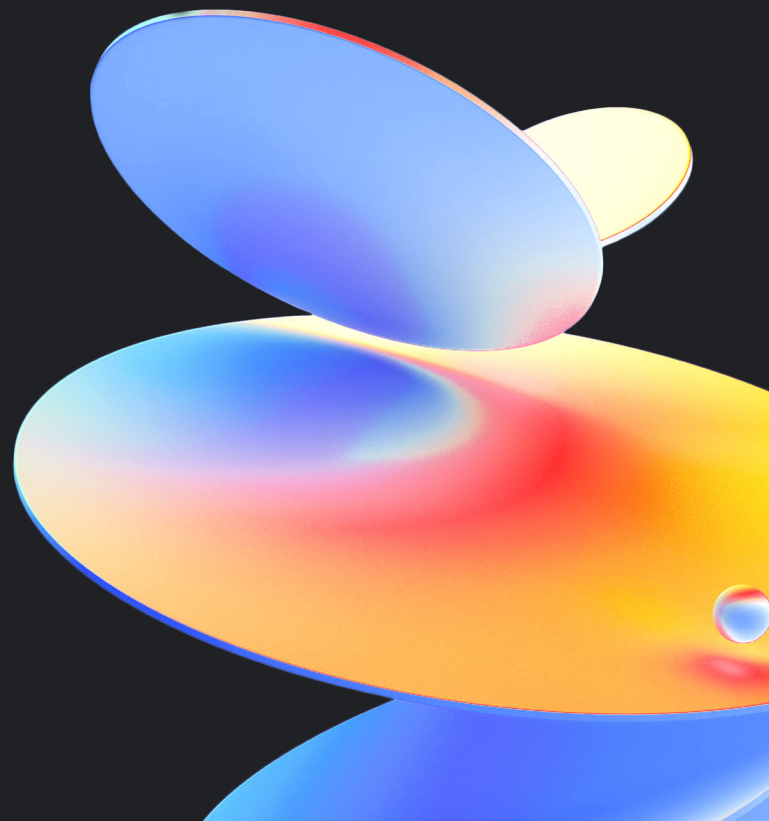
A quick primer on generative AI

"

AI has the potential to revolutionize the way the public sector operates, serves its missions, and supports its citizens. The ability for government leaders to use this technology responsibly and empower employees to use it hinges on imagining a new way of working and the ability to manage and respond to change."

Karen Dahut
CEO, Google Public Sector

Google

Every day, people in your organization spend time and energy digging for information to make decisions, serve customers, and move your mission forward. Informed decisions require information, and collecting the right inputs can take time.

Suppose you need to know how grant funding is leading to better outcomes.

Or you want to see the trends in public funding and how often projects are achieving on-time, on-budget milestones. Information like this exists somewhere in your organization, often across multiple sources. You need it to decide the next step — and, to obtain it, experts need to be convened, research completed, and the information compiled and synthesized. If you have a follow-up question, the whole process might have to start again.

From agency leaders digging into strategic trends to public health officials building outbound marketing campaigns or new employees with benefits questions — everyone in your organization can relate to this frustration. But that's changing.

Imagine giving each person in your organization not only a personal assistant, but an *expert* in every piece of data relevant to their job and, indeed, potentially every piece of data across your entire organization. With an assistant like that, impatient moments of indecision could dwindle. Everyone would be empowered to spend less time waiting and more time doing.

With generative AI, this is achievable. And it's just one example of the many disruptions this technology has unlocked always-on coding collaborators. Brainstorming assistants to draft and iterate content. Personalized self-education on any subject. Human-like, real-time interactions with constituents wherever they need you, for whatever reason. Generative AI is all these things and more. In time, it will affect almost every aspect of every agency mission and constituent touchpoint.

The transition will pose challenges. Generative AI applications are powered by foundation models. Foundation models are trained on vast amounts of content. For example, Large Language Models (LLMs) are one type of foundation model, trained on text or language. Other multi-modal types can be trained on images/photos, video, music, or software code. But access to models alone won't position your organization for success.

Google
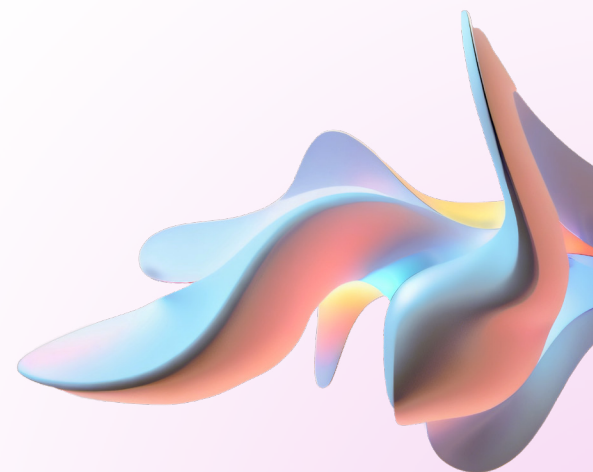
Foundation models are best thought of as probability engines that can be nudged and shaped by human input.

Because they are probabilistic, they are fundamentally different from traditional software paradigms. When today's apps need to look up information, they use deterministic functions to query a database. In contrast, foundation models use the patterns they learn during training and tuning to calculate the most probable output, such as the most likely answer to a question or an accurate caption for an image.

Since they're not confined by rows and columns in a database, foundation models are extremely powerful. They're often capable of performing many downstream tasks — such as Q&A, summarization, or open-ended content generation — with little or no additional data or tuning. Yet they can also be expensive to train and run, prone to inaccurate outputs, and difficult to work with.[3]

For these reasons, generative applications aren't reducible to generative models. Your intelligent applications will need to mix probabilistic foundation models with traditional, deterministic (in other words, constrained) programming. Deterministic models are constrained to what can flow out of them, and are limited by endless options that have to be pre-established.

Traditional AI is purpose-built for the task at hand. It is all about optimizing and tweaking existing processes such as forecasting specific patterns that had been pre-determined by humans. It's why traditional AI could be used to automate discrete, standardized processes in specific areas, such as customer service.

In comparison, generative AI models have emergent capabilities where they can perform multiple tasks, even if they were not explicitly trained for them during the instruction tuning phase. It is this multitask ability plus the freedom afforded by the prompting interface that allow these models to perform in a wide variety of use cases.

3. How to maximize your generative AI investments with cloud FinOps

Core capabilities and applications

At the core, generative AI capabilities can be grouped into the following categories:

✏️

Creation

Bringing your thoughts and visions to life

📋

Summarization

Condense and summarize large quantities of knowledge into an understandable format

🔍

Discovery

Help citizens, employees, and customers find what they're looking for

📊

Automation

Automate services across multiple channels

And it tends to excel in four applications:

Chat

It's no coincidence that generative AI has gained rapid popularity and adoption through simple chat interfaces. Chat is a natural way to interact with powerful generative AI models. You can use it to improve customer interactions, enhance product capabilities, train employees, and more.

Generate Content

The ability to generate high-quality text, images, speech, and code has enormous potential. Whether it's speeding up processes or helping employees turn ideas into output faster, generative capabilities can be deployed into products, tools, and workflows.

Search

By combining generative AI capabilities with search, you can anchor on a knowledge base — either internal or external — for more tailored and targeted interactions. Using generative AI for search can help eliminate hallucinations by sourcing information from a factual knowledge base.

Associative reasoning

This is the ability to suggest associations in information based on context, frequency, or proximity. For example, generative AI could identify the three most common reasons that a call center interaction ends negatively by parsing the large amounts of transcribed conversations.

The value of generative AI in public sector

To enable your foundation models to get very smart, very fast, pick one functional domain in your business and experiment in that area.

By clustering use cases into a domain, you can start with one and then, when that's working, you can naturally expand to your second, third, and fourth use case within the same domain. The more you throw at a model, the smarter it gets.

Take constituent-facing services, for example. First, let's say you give your call center agents a generative AI tool with a conversational interface to use when taking phone calls. Constituents might call in and say, "How do I apply for this benefit?." Or "I lost my passcode." Or "How do I get a Real ID?" Your agents can answer these questions naturally, querying the generative AI interface to provide an answer in the same language the caller is using.

Then, you can look at the aggregate of those queries and ask generative AI, "What questions are being asked most frequently? What's our response time? What responses do we provide?" Now, you're moving from just answering constituents' questions to summarizing their data.

Third, using this summarization, you can prompt generative AI to take the most frequently asked questions and compare them to the FAQs on your website. You can then ask it to generate answers to any questions not answered on your website, so you can publish them.

In this example, generative AI solves three use cases that enhance the constituent services domain — answer the question, summarize what's being asked, and produce answers to these questions. With each use case, the model gets smarter.
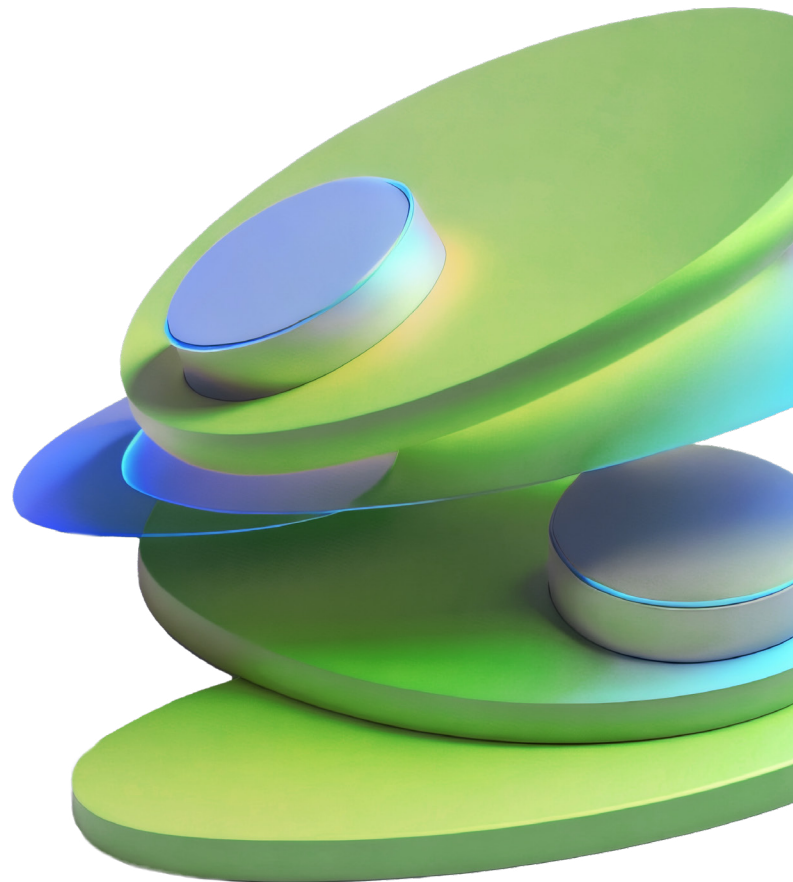
How to launch your first use case: Day 1- 30

We have identified the following steps to provide leaders with a change management guide to take a fast, easy, and low-risk approach to start your organization's generative AI journey. We've included KPIs to help you demonstrate impact to your leadership, foundational processes to scale across domains, and recommendations to allow your teams to experiment in a safe, internal setting so they can gain experience with the new technology.

In some cases these steps will have heavy involvement of your service partners if your organization currently relies on outside help to establish new IT systems. In these cases it is imperative to get consensus and buy-in from all parties that this is an important strategic initiative and everyone is working toward the same goals. We've seen great success in public/private partnerships in similar spaces and believe they could be workable models in the generative AI space.

Step 1

Identify a specific domain

Step 2

Select a persona

Step 3

Determine data sources the persona needs to be productive

Step 4

Create a small tiger team

Step 5

Define your intentions, objectives, and the output you are trying to achieve

Step 6

Design prompts with the tiger team

Step 7

Build a user experience (UX) and user interface (UI)

Step 8

Expand access to additional individuals

Step 9

Build a Language Model (LM) operations plan

Step 10

Expand usage to additional use cases within the same domain

Google

Identify
a specific domain

Typically, organizations in the public sector find three specific areas in which to leverage generative AI. Constituent services, empowering employees to be more productive, and deriving expertise from a generative model.

Key questions to consider:

In which areas of the business do employees spend significant time on repetitive tasks?

_____

Is there a process or part of a role that is, in effect, already standardized
(e.g. having to do a particular step or answer a question every time)?

_____

Where do employees get stuck in the creative process
(e.g. writer's block or creative block)?

_____

Is there an area of the organization where employees consistently have to search for existing information using internal knowledge bases and/or external search engines?

_____

Which areas of the business offer the lowest-risk environment for initial use cases?

_____

Would an incorrect response or hallucination cause harm?

_____

Do you have a large corpora of data that you want to activate, or make more useful?

Google

Select a persona

Determine which job category or function within the chosen domain you want to make more productive.

When selecting a persona, consider these three factors:

Look at job roles that are hard to retain and hard to hire.

Such roles are often repetitive and offer little career advancement. Automating these tasks can free up employees to focus on more strategic work. This leads to greater job satisfaction and higher employee retention rates.

Find opportunities to automate repetitive, tedious tasks for staff and constituents.

Making public services accessible to everyone is a founding principle of democracy. In 2021 the White House issued an executive order that "we must use technology to modernize Government and implement services that are simple to use, accessible, equitable, protective, transparent, and responsive for all people of the United States."[4] Reducing the paperwork required for accessing and processing government benefits saves time, money, and resources for both staff and constituents.

Look to places that often require outside resources to fit the need.

Many agencies have to procure expensive outside resources in order to keep the organization moving. Examples include translation services, call centers, or specialty software knowledge. These can often be great starting points to augment and assist with generative AI.

4. White House OMB Executive Order December 13, 2021

Google

Determine the data sources the persona needs to be productive

Identify and gather the relevant data required for the chosen persona. This data will be used to train the generative AI model, whose success depends upon data that has been honed for the specific organization or domain-level problem it is trying to solve.

For example, let's say you select a public benefits adjudicator role, a government individual who is responsible for determining what level of benefit a constituent is due based on their application. Usually in these cases, the constituent has entered their information either via an online form or via paper processes, and has attached multiple pieces of documentation. That application is likely "scored" by an internal system that attempts to understand what confidence the system has in their eligibility for that benefit.

This adjudicator role requires an individual to search against multiple data sources to identify the individual and their previous history, creating multiple correspondences to validate that history with external businesses or agencies, a CRM to track all correspondence, tasks, and workflows etc. All of these touch points need to be identified and looked into for applicability with the AI model.

"Google Public Sector is bringing the Google magic – evolving AI cloud-technologies, security, and scale – to the mission of government to transform the way citizens experience public services and the way agencies operate."

Leigh Palmer

Vice President of Technology Strategy & Delivery, Google Public Sector

Google

By starting off with the right data to feed into and fine-tune models, your organization will be able to:

Mitigate hallucinations

AI models are trained to give users what they want, which means they occasionally spit out made-up answers that sound convincing and can be hard to spot. To avoid this problem, you can ground responses in specific data rather than relying on the Large Language Model (LLM) alone.[5]

Enhance the explainability of AI

Generative AI models can be complex, and the 'thinking' that an algorithm uses to produce an output isn't always clear. Explainable AI is like a sliding scale — there are degrees to which you can explain or reliably steer an LLM's behavior.

5. Vertex AI Embeddings for Text: Grounding LLMs made easy

Create a small tiger team

Assemble a tiger team comprising an individual from each category.

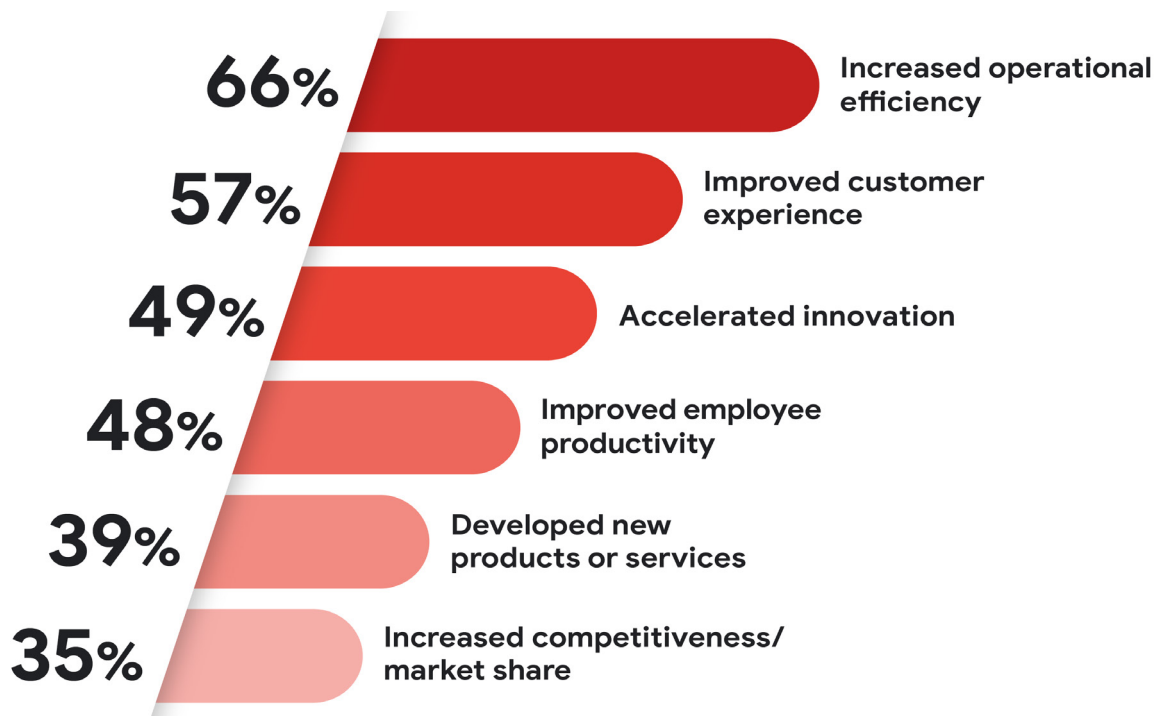|		
	Mission or key stakeholder	Responsible for detailing the job requirements, workflows, challenges, and needs of the day-to-day tasks executed by the chosen persona.
	Prompt engineer	Responsible for translating the business persona's needs, actions, and output into prompts for the generative AI model(s).
	IT operations lead	Responsible for working with the low-code no-code generative models and identifying the right production applications.
	Data owner	Responsible for ensuring that the data used to train the generative AI model is accurate, complete and up-to-date.

Google

Define your intentions, objectives, and the output you are trying to achieve

Ensure you have a human in the loop to oversee the first use cases and provide oversight.

Keep in mind, the value of a generative AI project can come from a number of sources. There's direct business value, incremental value of generative AI over legacy systems or traditional AI/ML, and the forecasted value of capabilities once scaled to other use cases.

Consider the following outcomes, which other organizations have experienced after adopting AI.[6]

Outcomes organizations report after adopting AI

66% Increased operational efficiency

57% Improved customer experience

49% Accelerated innovation

48% Improved employee productivity

39% Developed new products or services

35% Increased competitiveness/market share

6. Google Cloud Gen AI Benchmarking Study, July 2023

Google

Design prompts together with the tiger team

Work collaboratively with the tiger team to design prompts that will guide the generative AI model's responses.[7]

A prompt is simply a question or the command you give the model. By building prompts with your small tiger team — which has expertise in the organizational need, the AI model(s), tuning, and application integration — and referencing prompt samples[8], you will be able to move quickly.

7. Introduction to prompt design
8. Generative AI prompt samples

Build a user experience (UX) and user interface (UI)

Create a user-friendly experience and interface that will run the generative AI model in production for the chosen persona's use case.

Here are some tips to keep in mind:

Utilize low-code/no-code frameworks where available to simplify the process.

Keep the interface and design simple.

Begin with a persona selection screen, which allows users to choose the personality they want the generated text, images, or output to be stylized. For example, options could include "formal," "casual," "technical," or "creative."

Create a logical and intuitive user flow that guides users through the AI model's functionality. Ensure that the interface design aligns with the expected user journey.

Ensure the UI/UX is responsive and accessible across different devices and screen sizes, including mobile phones, tablets, and desktops.

Google

Expand access to additional individuals

Once you are getting acceptable results from tuning, invite 2-3 other individuals within the chosen persona to start using the model.

Continue testing, measuring, and tuning with this group until you are getting consistent, quality outputs, then expand usage to 5-10 more people within the chosen persona and continue fine-tuning the process.

With each new individual, make sure you understand the different ways each user interacts with the generative AI model. To do so, conduct user interviews, surveys, or workshops to gather insights into user preferences, pain points, and desired functionalities for the AI model's interaction.

Build a Language Model (LM) operations plan

Develop a plan for productionizing and monitoring the AI model's output to ensure it functions effectively and safely.

If you are working in a Public/Private partnership model, make sure to identify who owns which areas of this operation plan to ensure success.

Key questions to ask your management team:

Can we quickly evaluate and experiment with generative AI?

_____

What controls do we need in place during evaluation and experimentation?

_____

How are we measuring impact? Do we have targeted goals and frequent checkpoints to ensure progress?

_____

Do we have a mechanism for continuous improvement? Are we able to assess, evaluate, and re-engage to go deeper within existing use cases or expand to more use cases?

Google

An LM operations plan should include, but is not limited to:

Infrastructure setup

Prepare the necessary infrastructure for model deployment, including scalable compute resources and storage. Set up a version control system to manage model versions effectively.

Deployment and monitoring

Generative AI models can be complex, and the 'thinking' that an algorithm uses to produce an output isn't always clear. Explainable AI is like a sliding scale — there are degrees to which you can explain or reliably steer an LLM's behavior.

Create a harness to capture output and quality

Generative AI models can be complex, and the 'thinking' that an algorithm uses to produce an output isn't always clear. Explainable AI is like a sliding scale — there are degrees to which you can explain or reliably steer an LLM's behavior.

Continuous performance improvement and model updates

Performance – which is defined by quality and latency – requires updates to the model to incorporate the latest research advancements and improvements. Conduct A/B testing to evaluate the impact of model updates on safety and effectiveness.

Security and compliance

Ensure the entire system is secure, with appropriate access controls and encryption mechanisms to protect sensitive data. Comply with relevant regulations and develop AI governance principles that support your organization.

Human-in-the-loop oversight

Set up a human-in-the-loop process to review and moderate generated content, especially in sensitive or high-risk applications. Develop a feedback loop to continually improve the model's safety and effectiveness based on human moderation.

Incident response and remediation

Develop an incident response plan to handle potential safety breaches or issues promptly and effectively.

Regular evaluation for expansion

Establish a regular evaluation cadence to assess the quality of the AI-generated output and a plan for further expansion to other areas within the same domain.

Expand usage to additional use cases within the same domain

Repeat the process for 2-3 other use cases within the same domain. For example, in the beginning of this chapter, we explained how an organization could start with one generative AI use case and naturally expand into three use cases that could all enhance the constituent service domain. First, it helped call center agents answer the question, then, second, it summarized the frequently asked questions agents received over the phone, and third, it produced answers to these questions that could be posted online as written FAQs.

With each use case added to the model, the model itself became more accurate in the domain.

Looking ahead at days ━━ 60–90 ━

Once you are ready to extend your use case to external users and/or third-party data, methods and tactics you can use to scale quickly and safely include:

01

Host a hackathon

Performance – which is defined by quality and latency – requires updates to the model to incorporate the latest research advancements and improvements. Conduct A/B testing to evaluate the impact of model updates on safety and effectiveness.

02

Spread the word through innovation challenges or accelerators

Ensure the entire system is secure, with appropriate access controls and encryption mechanisms to protect sensitive data. Comply with relevant regulations and develop AI governance principles that support your organization.

03

Encourage civic engagement

Open-sourcing your data sets and helping the public understand how to access and utilize them can create new avenues of utilization for these valuable data sets.[9]

04

Create a center of excellence

Excitement about new technologies can often lead to widespread use. A center of excellence in models, tuning, and application integration can help to standardize processes (including governance), share knowledge, and ultimately drive innovation.

05

Bring in partners

In addition to sharing their expertise, partners can consult on mission value and technical implementations, provide training, and even work side-by-side with your teams to transfer knowledge as they build out your implementation.

9. The Keyword: Data Commons and AI

Google

A word of advice: Do not wait to gain experience

Even in the private sector, over 40% of organizations take over 6 months to bring models from concept to live production, and 80% take over 3 months.[10]

Given the complexities and challenges of data and AI, it can be easy to get bogged down. Yet as time and costs stack up, the struggle to get use cases into production leads to the fear that positive ROI will never be realized from your AI investments.

This is why it's so critical to move quickly, experiment, and gain experience with production AI. It's never been more important for the public sector to understand a new technology and how to master it. Additionally, your agencies need to lead from the front to ensure ethical and responsible use of these AI techniques.

The steps in this 30-day plan will help leaders lay the groundwork in a low-risk environment by experimenting internally with generative AI in your organization. The strength of your team and their ability to move quickly toward your mission goals will ultimately determine the success of your AI investment — no matter whether your investment is low, medium, or high; or whether you build or buy (or both).

In the next section, we will cover measurement and recommend a list of commonly used KPIs to measure the impact of your generative AI use cases.

KPIs that measure the impact of generative AI

When evaluating projects, consider the feasibility, actionability, affordability, anticipated business value, and ultimate return on investment of each generative AI project.[11]

Like any technology investment, you need to prove its worth. Embed ROI measures into every use case and project, and establish KPIs to keep a pulse on progress along the way.

Consider using these commonly used generative AI KPIs to measure and report on the value of generative AI to your organization, board members, and stakeholders. These KPIs apply to generative AI use cases across various domains and industries.

11. How to maximize your generative AI investments with cloud FinOps

Google

Accuracy

Measure the accuracy of the Measure the accuracy of the generative AI model in producing relevant and correct outputs. This can be quantified using metrics such as precision, recall, F1 score, or mean squared error, depending on the nature of the use case.

Productivity

Assess the impact of generative AI on the productivity of the target persona or department. This could include metrics like the number of tasks completed per unit of time, response time, or reduction in manual effort required.

Customer satisfaction

If the generative AI use case involves customer-facing applications, use customer satisfaction surveys or feedback to gauge how well the AI system meets customer needs and expectations.

Cost savings

Measure the cost savings achieved through the use of generative AI. This may involve comparing the costs of employing the AI system to the expenses associated with traditional manual processes or outsourcing.

Turnaround time

Evaluate the time taken for the generative AI model to generate responses or outputs compared to traditional methods. Faster turnaround times can lead to increased efficiency and improved customer experience.

Quality of output

Assess the quality of the generative AI outputs against predefined criteria. This can be done through manual reviewor automated quality checks, depending on the use case.

Error rate

Quantify the rate at which the generative AI model produces incorrect or undesirable outputs. Minimizing error rates is crucial for maintaining accuracy and reliability.

Business impact

Identify specific business metrics that are directly impacted by the generative AI use case, such as increased sales, reduced customer complaints, or improved employee retention.

Training time and cost

Measure the time and resources required to train and fine-tune the generative AI model. Efficient training processes can lead to faster implementation and quicker time-to-value.

Human-in-the-loop metrics

If human intervention is involved in the generative AI process, track metrics related to the efficiency and effectiveness of human oversight.

Scalability

Assess how well the generative AI model scales to accommodate increased usage or higher demands. Scalability is essential for long-term success.

Regulatory compliance

For sensitive domains like healthcare or finance, monitor how well the generative AI system adheres to relevant regulatory requirements and data privacy standards.

# The step-by-step guide to getting started with generative AI

Generative AI is more than a bright shiny object —
it is an entirely new value stream for public sector
leaders. Already, government agencies and
defense organizations are using generative AI
to solve some of their most common and
time-intensive problems.

For instance, generative AI can improve the efficacy of virtual agents by increasing the size and scope of data that the models are trained on, offering even more insightful and natural conversations to constituents in real time

Government agencies can use generative AI to extract information and automate paper-based processing, and create systems that automatically generate summaries and reports on government activities. These reports could help increase transparency and accountability and improve decisions around how to best allocate resources.

Further, generative AI can help enhance efficiencies in operations by identifying patterns in data, predicting trends, and making recommendations.

In this section, you will learn how public sector leaders are applying generative AI to the top use cases across domains to unlock new value, rewrite processes, and advance their missions in faster time frames, and at a lower cost.
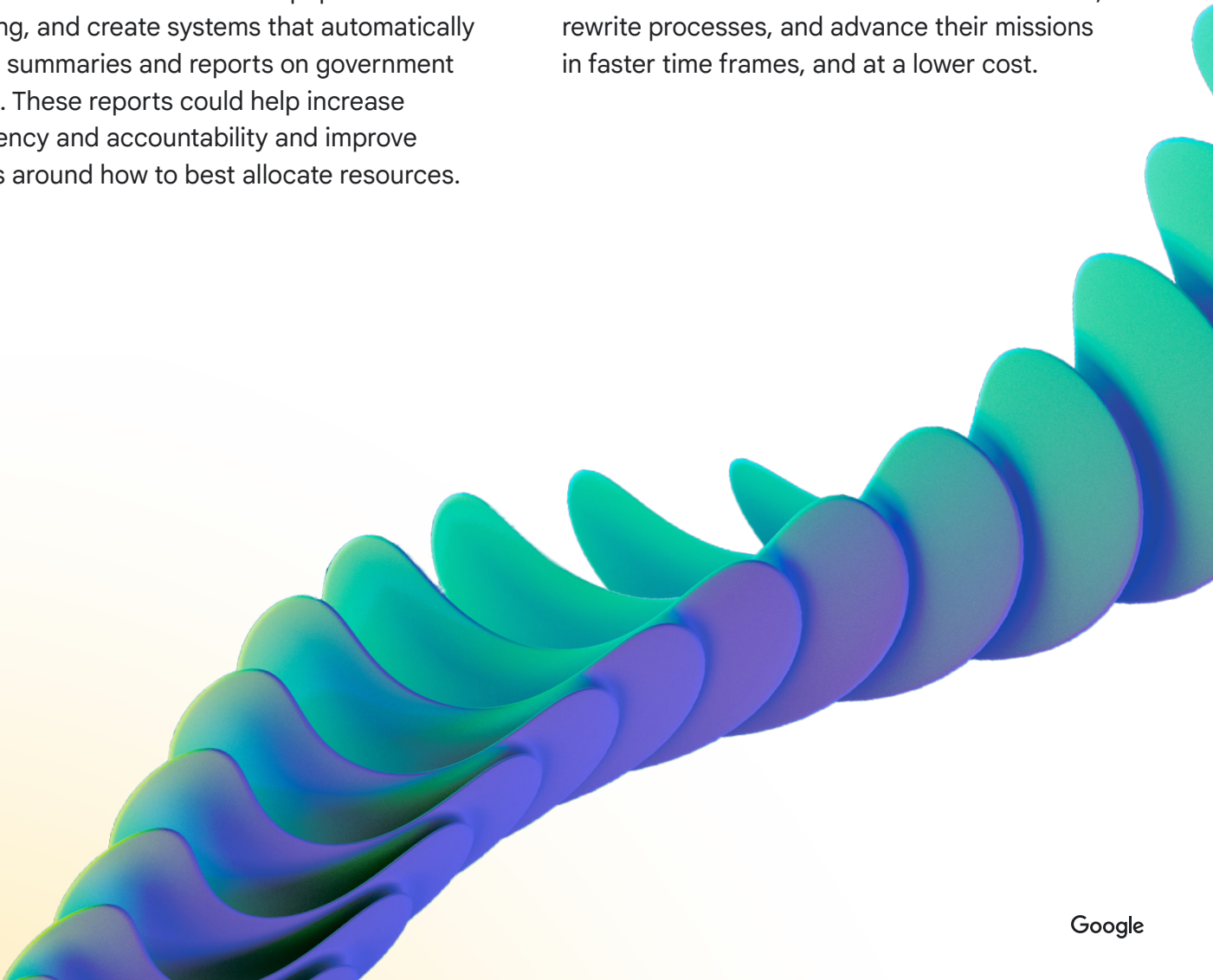
Google

Generative AI across different segments

Google Cloud provides a secure platform to support government operations and constituent experiences. First, it helps organizations positively transform the constituent experience – allowing you to control your site, your data, all through a virtual agent, powered by gen AI, that communicates, translates, and takes action. Second, it provides employees with additional superpowers: as if an AI-powered mentor is sitting beside the employee helping them search, summarize, and understand information.

Third, it enables expertise on-demand. Imagine having access to security experts, data analytics experts, translators, and software developers at all times without the human premium.

To accelerate your path to value, familiarize yourself with the industry-specific use cases being adopted today. See how the leaders are realizing value in their organizations, and get inspiration for your own.

Explore them all or jump straight to your sector:

Government →

Health & Human Services →

Labor →

Transportation →

Education →

77%

of government decision-makers would like their organizations to adopt artificial intelligence (AI) more aggressively and feel their employees are prepared to do so.[12]

12. KPMG study on AI in government

Government

Priority use cases

Improve agency operations
Vertex AI Search for Government

Conducting enterprise-level searching is a complex and challenging undertaking. It can be difficult to implement and maintain an effective search system due to a number of factors, including data silos and incomplete data, complex search queries and relevance issues, integration with backend systems and databases, security and privacy concerns, and cost and scalability.

Every government agency tends to be document rich and insight poor. Vertex AI Search empowers your employees to be able to summarize and derive insights from all the data your organization has accumulated. Agencies whose employees need to be able to look through hundreds of pages of documentation to find the correct answer to policy questions can now enter a simple question into a search box and receive a summarized answer grounded in the documentation provided.

Vertex AI Search uses the same semantic search technology as Google Cloud Search to search across your Google Workspace data as well as third-party data sources, but it gives you more control over the way the search engine is configured. It uses Google's search technology to deliver relevant results, and it can be customized to meet the specific needs of your organization.

Insight

92%

of U.S. citizens report that "improved digital services would positively impact their view of government."[13]

13. Accenture Survey Shows U.S. Citizens Want More Digital Services from Their Government, December 2015

Government

Priority use cases

Enhance constituent services
Vertex AI Conversation for Government

Constituents' expectations of how they can get information from government agencies have never been higher. Consumer applications have gotten to the point where customers can get access to their data whenever and wherever they are. Governments have historically struggled to offer this same level of self-service due to technical constraints. Vertex AI Conversation offers a single tool that can 10x the way constituents interact with their governments.

Vertex AI Conversation is a rapidly growing technology with the potential to revolutionize the way governments interact with their constituents. This new branch of generative AI gives you the chance to deploy a virtual agent through a variety of mechanisms that knows your agency data and can answer constituent questions and offer a 24/7 multilingual agent that can help constituents get the answers they need in real time.

Google has a suite of tools and services to help public sector organizations leverage its benefits quickly:

- Dialogflow for building and deploying conversational AI agents

- Agent Assist for integrating conversational AI into existing contact center applications

- Contact Center AI Platform for building and deploying conversational AI applications for contact centers
- Speech-to-Text for transcribing customer conversations, or creating text-based chatbots

- Text-to-Speech for creating voice-based chatbots, or generating audio content for constituent engagement applications

Supercharge your workforce - Creative Assistance

With Duet AI for Google Workspace, government employees are armed with tools that expedite the content development process. This integrated creative assistant helps workers save time and maximize efficiencies through helping to visualize, organize and accelerate workflows. For instance, agencies can streamline the recruitment process, creating and refining job description information. Additionally, Duet AI can be used to generate reports, presentations, and other documents, as well as to create and translate content. This can be a valuable tool for a variety of government areas and missions, such as policy development and grant writing, where Duet AI can help to quickly generate and refine content.

Government

Priority use cases

Improve accessibility
Multilingual Chat

Google Cloud offers a number of solutions to help government agencies interact with constituents across language barriers and cultural difference, including:

- Dialogflow CX: for building conversational interfaces for your apps and websites in over 100 languages.

- Cloud Translation: for translating text and speech in real time so that users can communicate with your chat bot in over 100 languages, at any time.

- Business Messages: for building chatbots within Google Chat in over 20 languages to help agencies interact with their communities.

Insight

74%

of state CIOs said that delivering a better online experience for citizens was the biggest driver to expanding digital services.[14]

Google Cloud's Natural Language API

Gaining insight from free text fields can be a significant challenge for agencies. Whether the data sets are call centers logs, comment fields in surveys, or internet content like news stories or social media posts, there is a consistent need to gain understanding on the underlying intents and entities. Natural Language API gives agencies the ability to more quickly and easily derive these insights in a scalable and automated fashion.

- Sentiment analysis: understand the overall opinion, feeling, or attitude expressed in a block of text. For example, a government agency could use sentiment analysis to gauge public opinion on a new policy or program.

- Entity recognition: identify and extract information from documents. For instance, using entity recognition, government agencies can automate the processing of claims and government documents, saving time and improving the accuracy of data entry.

- Text classification: automatically classify documents in 700+ predefined categories.

- This can help organizations improve the efficiency and effectiveness of their operations.

14. Driving digital acceleration: The 2021 State CIO Survey, NASCIO

Government

Deep-dive

Increasingly, constituents have come to expect from state and local governments the interactive, always-on digital-first experience that they are used to as consumers. Yet government agencies may not be able to adapt and scale quickly enough to meet the level of demand and evolving need for citizen services. In addition to security and data governance considerations, government agencies must also address accessibility and equity concerns to reach traditionally underserved communities. Democratizing access to services ensures that everyone can renew their drivers' licenses, check their polling locations, and review school-zone requirements. Visionary agency leaders are already using AI and predictive analytics to improve services like these and better allocate resources.

As the volume and complexity of constituent service demands continue to increase for state and local governments nationwide, agencies are turning to AI to help minimize wait times and provide vital information to constituents. Tools like Dialogflow can enable this process by providing a state-of-the-art virtual agent experience that features lifelike, conversational AI.

Real-world examples

Minnesota expands 24/7 access to its Department of Motor Vehicles (DMV)

In Minnesota, over 10% of the population does not speak English, which can create challenges for certain groups, such as resettled refugees, in meeting driver vehicle requirements. The Minnesota Department of Public Safety (MN DPS) recognized that this language barrier hindered these groups' ability to obtain mandatory insurance and REAL ID-compliant driver's licenses. This, in turn, restricted their mobility and access to services.

In collaboration with Google Public Sector, MN DPS introduced Google Cloud Translation AI to their services. This provides constituents with access in Spanish, Somali, and Hmong languages, as well as English. This effort promotes equitable services for underserved Minnesotans and contributes to the creation of a Minnesota where everyone has an opportunity to thrive.[15]

City government reaches communities in their own languages with the help of AI

The city of Dearborn, MI implemented a call center[16] so residents can access city services in multiple languages, anywhere, at any time. Previously, many city services required physically going to a government office during business hours, which often meant taking time off work and waiting in line. Now, residents can request and process critical actions—like applying for permits and licenses—faster and online. Virtual agents on the website deliver personalized information and interactions in three languages (Arabic, English, and Spanish) with chat and voice, powered by Google Cloud Contact Center AI (CCAI). The city website and documents will be translation-enabled by Translation Hub and Document AI, which will help expand the reach and delivery of vital government information and services.

15. The City of Dearborn & Minnesota's Department of Public Safety digitally transform with Google Cloud, 2023
16. Dearborn transforms its digital services, becoming a model for American cities

AI can have a significant impact across mission areas.

The public sector serves a variety of mission areas and objectives where AI can have a significant impact.

To start your imagination here are a few examples:

_____

Reimagine the healthcare journey by providing a multilingual, user-friendly portal that connects patients and visitors. This portal can provide enterprise search for providers or medical centers, as well as the ability to make appointments and give consent to share health data.

_____

Transform threat detection, investigation, and response for cyber defenders. This can help to simplify search, complex data analysis, and threat |detection engineering, easing the burden on security analysts.

_____

Leverage data as a strategic asset to help inform fleet management. This can include real-time tracking, document processing, and route optimization.

These are just a few examples of the many ways that AI can be used to improve government operations. By using AI, government agencies can become more efficient, effective, and responsive to the needs of their constituents.

Health and Human Services

Priority use cases

Conducting enterprise-level searching is a complex and challenging undertaking. It can be difficult to implement and maintain an effective search system due to a number of factors, including data silos and incomplete data, complex search queries and relevance issues, integration with backend systems and databases, security and privacy concerns, and cost and scalability.

Every government agency tends to be document rich and insight poor. Vertex AI Search empowers your employees to be able to summarize and derive insights from all the data your organization has accumulated. Agencies whose employees need to be able to look through hundreds of pages of documentation to find the correct answer to policy questions can now enter a simple question into a search box and receive a summarized answer grounded in the documentation provided.

Vertex AI Search uses the same semantic search technology as Google Cloud Search to search across your Google Workspace data as well as third-party data sources, but it gives you more control over the way the search engine is configured. It uses Google's search technology to deliver relevant results, and it can be customized to meet the specific needs of your organization.

Insight

98%

of the 500 health care organizations in the 2021 Optum Survey on AI in Healthcare either have an AI strategy or are planning one.[17]

17. 2021 Optum Survey on AI in Healthcare

Health and Human Services

Priority use case

Provider Update Portal

Provide a unified search engine for constituents to find health-related resources in an easy way, including providing accurate and up-to-date provider and community-based information.

Master Person Index for Health Records

Provide a way to integrate disparate data sets across multiple systems and platforms to build a robust health analytics platform that enables self-service, provides a better view of social determinants to improve health outcomes.

Medicaid Intelligent Redetermination Agents

Use intelligent virtual agents to enable self-service to help constituents determine if they are eligible for Medicaid.

Deep-dive

Harness data to improve patient outcomes while protecting privacy

Generative AI is opening up faster and more conversational ways to discover trends in healthcare data. It can summarize information from large volumes of content, and automate time-consuming and expensive processes. It can help clinicians access and understand the data they need to work faster and smarter, which ultimately impacts patient care. It is possible to do all of this while maintaining full control of your data, and complying with various regulations and data sovereignty laws.

Medicine, after all, is about caring for people. Over the past five years at Google Health, our research has shown that AI can augment a clinician's ability to detect breast cancer. It can help people better understand their skin conditions, and it can help researchers sequence genomes more accurately than ever.

Conversational AI systems in medicine like Med-PaLM 2—a large language model designed to provide high quality and authoritative answers to medical questions—are being used to explore how AI can help clinicians spend less time on contouring.[18] Much like Google Maps can detect buildings and roads using AI, technology could detect anomalies in CT scans, quickly outlining them for review by specialists.

18. The Check Up with Google Health, 2023

Google

Health and Human Services

Real-world examples

U.S. Department of Defense builds AI-powered microscope to advance cancer care

The DoD's Defense Innovation Unit (DIU) contracts with commercial companies to help the military integrate the latest technologies.

For several years they have been working with Google to develop an Augmented Reality Microscope (ARM) that uses AI to analyze and evaluate the severity and spread of cancer in a cell. This can help overburdened pathologists validate their diagnoses and offer the benefits of enhanced vision. So far, ARM is being trained to identify breast cancer, cervical cancer, prostate cancer, and mitosis, with promising results. By supplementing clinical expertise, AI can improve patient outcomes and mitigate workforce shortages.

HCA Healthcare transforms patient care with generative AI

According to a 2022 study in JAMA Internal Medicine, 58% of physicians said time spent on documentation limits the amount of time they can spend with patients. As part of a pilot program that began early in 2023, approximately 75 emergency room physicians at four HCA Healthcare hospitals started using Google's AI technology to quickly and more easily document key medical information from conversations during patient visits.[19] They deployed Vertex AI Search to improve the efficiency of clinical workflows, so clinicians and researchers can find the information they need, and ultimately improve patient outcomes. Vertex AI Search unifies data across dispersed documents, databases, and intranets, making it easier to search, analyze, and identify the most relevant results.

> "Generative AI and other new technologies are helping us transform the ways teams interact, create better workflows, and have the right team, at the right time, empowered with the information they need for our patients."

Michael J. Schlosser
MD, MBA, FAANS, SVP,
Care Transformation and Innovation, HCA Healthcare

19. HCA Healthcare Collaborates With Google Cloud to Bring Generative AI to Hospitals

Labor

Department of Labor (DOL) agencies face a number of core challenges in serving unemployment needs, including the high volume of claims, complex eligibility requirements, fraudulent claims, and lack of resources. These challenges can make it difficult for DOL agencies to serve the needs of unemployed workers. Google Cloud AI can help DOL agencies to address these challenges by providing them with scalable, intelligent, and secure solutions. For example, Google Cloud AI can be used to automate tasks, provide personalized services, detect fraud, and improve decision-making.

Insight

The GAO estimates that the amount of fraud in unemployment insurance programs during the COVID-19 pandemic was likely between $100 billion and $315 billion.[20]

Priority use cases

Fraud Analytics and Detection

Fraudulent claims can cost the DOL billions of dollars each year.[21] Traditionally, the DOL has relied on manual methods to detect fraud, but these methods can be time-consuming and labor-intensive. AI-powered analytics can help the DOL to automate the fraud detection process and improve the accuracy of fraud detection.

By using AI, the DOL can reduce the costs associated with fraud and protect taxpayer dollars.

Adjudication Assistant

Provide adjudicators with access to relevant data through search engine capabilities to rapidly query and cross-reference data to expedite the time it takes to process claims. Additionally, by providing adjudicators with a comprehensive view of all relevant data, Adjudication Assistant can help to improve the accuracy of claim decisions.

Personalized Services

Assist with full claim review and generate suggested scripts for agents to help agents focus on more complex claims.

Agile Queries

Provide labor agencies with "agile queries" that leverage gen AI constructs (e.g., LLMs) to perform queries against data sets and respond to employees in a relevant and consistent format – enabling workers to make informed decisions quickly and efficiently.

20, 21. GAO, Unemployment insurance

Labor

Deep-dive

Increase constituent service capacity and speed

Unemployment programs face challenges with scalability, data, and complexity. Federal extended benefits programs, which are typically implemented during recessions, add complexity, making it difficult for governments to scale programs quickly and effectively. This can lead to delays in providing benefits to constituents. Additionally, the lack of meaningful reports and dashboards makes it difficult for governments to make informed decisions about how to best administer unemployment programs.

Contact centers are exception-based, with a majority of constituent interactions taking place via self-service. However, complex questions and populations with barriers are difficult to serve. Call center staff require 6-12 months of training to competency due to complexity. Adjudication is also challenging. It takes two years to become an adjudicator, which could be an entire recession.

Google Cloud's AI portfolio can help DOL agencies address these challenges by providing solutions for fraud prevention/detection, AI translation services paired with chatbot capabilities to improve constituent experience, generative AI to support "Fact Finders" to assist during adjudication process, and generative AI enhanced Call Center AI (CCAI) Agent Assistants.

Labor

Real-world examples

Wisconsin department improves services with AI

The Wisconsin Department of Workforce Development deployed Google Cloud AI and ML for predictive analytics.[22] As a result, it cleared its unemployment application backlog and was able to process an average of 157,000 claims each week, releasing most payments to constituents within two to three business days. This helps government employees save time and energy, while improving service delivery to the community.

Illinois agency accelerates claims with AI

During the pandemic, the Illinois Department of Employment Security (IDES) used Google Cloud's Contact Center AI to rapidly deploy virtual agents to help more than 1 million citizens who lost their jobs to file unemployment claims. Web and phone-based virtual agent systems were each deployed in less than two weeks, and handled 140,000 phone and web inquiries per day.

> "The automated virtual agents have acted like a force multiplier for IDES's support agents, in terms of processing and responding to unemployment benefits requests."
>
> **Jennifer Ricker,**
> former Acting Assistant Secretary,
> Illinois Department of Innovation & Technology.

22. Economic Recovery: Wisconsin Leans on Google Cloud to Better Serve Its Community During COVID-19 and Beyond

Google

Transportation

Transportation agencies at all levels of government have a critical mission: to provide safe, efficient, and accessible transportation for people and goods.

This mission is becoming increasingly complex and challenging, as transportation systems become more congested and interconnected. AI can help transportation agencies to address these challenges and make our transportation systems safer, more efficient, and more accessible for everyone.

Insight

The global market for AI in transportation is expected to grow from $3 billion in 2022 to $23 billion by 2023.[23]

23. Artificial Intelligence in Transportation Market

Priority use cases

Predicting traffic congestion and delays

AI can be used to analyze historical traffic data and traffic sensor data (including public transportation cameras) to predict traffic congestion and delays. This information can be used to develop strategies to reduce congestion and improve traffic flow. This prediction can be particularly useful in supporting emergency preparedness and evacuation planning.

Identifying and prioritizing maintenance needs

AI can be used to analyze asset condition data to identify roads, bridges, and other infrastructure that need maintenance. This information can be used to prioritize maintenance work and ensure that the transportation network is safe and reliable.

Promoting sustainable transportation

AI can be used to develop strategies to promote sustainable transportation, such as public transportation, biking, and walking. This can help to reduce traffic congestion and improve air quality.

Transportation

Deep-dive

Use AI with your data for more intelligent predictions and decisions

Unemployment programs face challenges with scalability, data, and complexity. Federal extended benefits programs, which are typically implemented during recessions, add complexity, making it difficult for governments to scale programs quickly and effectively. This can lead to delays in providing benefits to constituents. Additionally, the lack of meaningful reports and dashboards makes it difficult for governments to make informed decisions about how to best administer unemployment programs.

Contact centers are exception-based, with a majority of constituent interactions taking place via self-service. However, complex questions and populations with barriers are difficult to serve. Call center staff require 6-12 months of training to competency due to complexity. Adjudication is also challenging. It takes two years to become an adjudicator, which could be an entire recession.

Google Cloud's AI portfolio can help DOL agencies address these challenges by providing solutions for fraud prevention/detection, AI translation services paired with chatbot capabilities.

to improve constituent experience, generative AI to support "Fact Finders" to assist during adjudication process, and generative AI enhanced Call Center AI (CCAI) Agent Assistants.

Transportation

Real-world examples

City of Memphis deploys AI to fix roads

The City of Memphis spends approximately 32,000 man-hours each year repairing potholes on its city streets. Historically, residents reported potholes and blighted properties by calling 311, or more recently by using the Memphis 311 app. However, these reports only covered about 20 percent of the problems — often the worst cases. Recognizing that potholes and vacant properties are often the most visible indicators of whether a city government is doing its job efficiently, Memphis Mayor Jim Strickland and CIO Mike Rodriguez began looking for ways they could apply technology to fix the problems.

The city government worked with Google Cloud and SpringML, a Google Cloud Premier Partner, to identify and fix potholes and determine areas of urban blight by analyzing high-resolution video footage with AI and ML.[24] The team projects 75% of potholes will be identified as a result, improving streets, communities, and the experiences of both residents and visitors.

Eagle County, Colorado transformed its emergency operations to better serve residents

Rapid, reliable coordination is essential during emergencies, as responders need to work quickly to minimize threats to civilians and damage to property. The Lake Christine Fire, which started in July 2018, created many challenges for Eagle County officials. Beginning outside Basalt, Colorado, the fire burned more than 12,000 acres throughout the region. Within this high-pressure scenario, Eagle County reinvented its emergency operations center with the help of Google Cloud tools.

County officials developed an interactive online fire map that could collect and analyze data, share it with community residents and staff in one accessible location, and track real-time conditions. During the wildfire people could see information about road closures, driving directions, and evacuation center locations. County officials could efficiently redirect people to other emergency pages right within the map. For potential floods following the fire, the county also exported its internal analysis information and created a second map that residents could use to see if their homes fell within a flood risk zone. With this map, at-risk homeowners could easily learn about insurance subsidies and other flood resources.

This solution transformed the Eagle County Emergency Operations Center, which had previously relied on a legacy workflow environment with specialized industry features. In times of crisis, the system struggled to support collaborative work.

24. City of Memphis: Detecting potholes for better citizen experiences

Education

Personalized learning has been an elusive goal in education for the past decade, but until recently, the tools to make this easy and cost-effective for institutions and EdTech companies were not available.

With AI solutions advancing so quickly, we see this impact education in three big ways: deep personalization, extreme productivity, and expanded access. EdTech companies, schools districts, higher ed institutions, and researchers will explore new ways to teach, learn, and discover. Organizations of any size will be able to make learning more personal, safer, and accessible.

Priority use cases

AI Tutors and Mentors

Create personalized course content that is available 24/7 to meet the unique needs of learners

Student/Operational Metrics

Unify student activity data across school ecosystems of educational tools, allowing for vertical integration across systems, and longitudinal insights across the learner's entire educational journey

Faculty/Instructional Design Support

Leverage Google gen AI services to a) generate syllabi and lesson plans that are tailored to the specific needs of a course and its students, b) personalize course material, and c) create simulations and virtual environments to teach students complex concepts

Alumni Engagement

Deploy Google gen AI services (e.g., Call Center AI) to engage with alumni to promote their universities, mentor students, and shape the future of their universities

Intelligent learner recommendations

Deliver personalized content at scale to each learner by auto-tagging metadata across various file types to enable smart cataloging and search

Content generation and localization

Make your content more accessible by translating between text and audio in over +130 languages using our pre-trained or your custom models

Marketing optimization

Automate your marketing strategy with real-time analytics and AI to score leads and engage with the right learner, educator, school, or workforce

Education

Deep-dive

Support student and faculty success with service automation and personalized content

The City of Memphis spends approximately 32,000 man-hours each year repairing potholes on its city streets. Historically, residents reported potholes and blighted properties by calling 311, or more recently by using the Memphis 311 app. However, these reports only covered about 20 percent of the problems — often the worst cases. Recognizing that potholes and vacant properties are often the most visible indicators of whether a city government is doing its job efficiently, Memphis Mayor Jim Strickland and CIO Mike Rodriguez began looking for ways they could apply technology to fix the problems.

The city government worked with Google Cloud and SpringML, a Google Cloud Premier Partner, to identify and fix potholes and determine areas of urban blight by analyzing high-resolution video footage with AI and ML.22 The team projects 75% of potholes will be identified as a result, improving streets, communities, and the experiences of both residents and visitors.

Insight

AI has the highest CAGR growth of 35% of all advanced technologies in education and its expected market size in 2025 will be $6.1B.[25]

Google

Education

Real-world examples

IBL Education launches generative AI-enabled mentors

With more than 6 years of experience in building open source and generative AI in education at scale, IBL Education provides enterprises and academic institutions with a platform to build, train, and securely customize large language models (LLMs) for interactive mentors using Google's Vertex AI.[26] The company leverages their strong analytics and AI foundation to personalize, translate, and create education content faster than before. This has been especially effective in improving accessibility, as automated translation and repurposing ensures those with visual and hearing impairments can interact with the content.

IBL Education allows users to create virtual mentors that provide personalized teaching, assess student knowledge, guide students through skills and learning paths, and offer robust learning analytics. Text-to-text and speech-to-speech interfaces can be offered over Slack, Discord, text message-based bots, web scripts, and LTI integrations.

"

AI Mentor has been a very effective tool for our users because it adapts to the needs of educators and organizations very quickly. We're seeing people use it for teaching assistance, marketing and administration, and a lot more. Vertex AI allows us to have many models for any variety of purposes."

Miguel Amigot II
Chief Technology Officer, IBL Education

26. IBL Education's GenAI-based chat mentor with Google

Google

Education

Real-world examples

Varsity Tutors uses AI to enhance individualized learning

Varsity Tutors is a Live Learning Platform designed by Nerdy Inc. to seamlessly connect learners to experts in more than 3,000 subjects. The platform uses AI to match students to tutors, factoring in over 100 variables affecting academic, social, and motivational outcomes. Tutors on the platform also use AI-driven adaptive assessments to quickly understand a student's grasp of a subject, identifying strengths and opportunities for growth at the concept level. Students can access personalized AI-generated quizzes, study guides, and assessments to practice and monitor progress. Varsity Tutors has developed over 100,000 practice problems, spanning 200 subjects across 4,000 skills, each with detailed answers and explanations. Additionally, students can enjoy continuous, 24/7, real-time support through an AI-based chat.

By building proprietary models on top of various foundational models on Vertex AI, Varsity Tutors makes learning even more personal and accessible. For example, Vertex AI makes it possible to combine text and AI image generation, which helps break down complex topics into powerful images to help students visualize a topic.

> The exciting prospect lies in AI's potential to revolutionize the learning landscape. By seamlessly integrating AI into every aspect of our platform, from crafting AI-generated lesson plans to intelligently matching experts with learners, we dramatically enhance our ability to customize the learning experience to an individual's unique needs, interests, and goals."

Chuck Cohn
Founder and CEO of Nerdy Inc.

Choosing the right solution for your organization

From tools and services with built-in AI/ML for non-technical users, to ready-to-use models that developers can easily implement, to advanced customizations led by data scientists and ML experts, generative AI implementations carry varying degrees of complexity.

It is important to consider total cost in ownership (TCO) when determining what is the right model for your organization. Considerations for TCO include size of the task (i.e. what business case you are trying to solve), size of the model, and performance (quality and latency). One of the fallacies we see made in these early days of generative AI is an over focus around individual model performance. The myriad of use cases and complexities addressed in this document show a need to identify the overall AI platform that your agency or organization can utilize most effectively. It's key to cut through the complexity and look at overall platform controls.

To manage this complexity, it's important to strike a balance between the capabilities you need to drive your use case and the level of investment you're willing to make. At this stage, you need to make at least two executive decisions.

Google

Build vs buy

Like any technology, as the capabilities of generative AI become more sophisticated, the layers of abstraction become simpler and easier to use — offering you a wide range of ways to interact.

In many cases, leaders find a mixed strategy works best. A platform-based approach offers the foundation needed to expand AI experience and expertise to deliver more value, more use cases and better performance. Yet, in many cases, you'll realize value faster by choosing solutions or managed services that fit your needs.

Key questions:

Are we capable of building, deploying, and training an LLM?
If not, where do we get the LLM from?

How core is this toolset / capability to our organization
(i.e. How directly is it tied to an agency's mission)?

What do we value more – gaining hands-on AI experience
or deploying a solution quickly?

Does building in-house give us a new unique service? How so?

What option creates impact through simplicity?

What's the fastest way to go live?

Do we have engineering / PM experience & core competencies needed to build it?

What's the cost of building vs. buying?

Google

Off-the shelf vs. customized vs. proprietary

There are multiple techniques to "tune" or customize a model. Prompt-tuning and reinforcement learning with human feedback (RLHF) are the fastest and most common methods customers are using to train foundation models.

Off-the-shelf

This is the easiest way to get started, either with pre-trained models with Google's Model Garden or with generative AI features embedded in your productivity tools. The first approach allows developers to quickly integrate generative AI into new and existing applications via simple APIs. The second approach allows individual users or teams to use generative AI embedded in their productivity tools to help them write or summarize emails, chats, and documents, organize data in spreadsheets, or generate custom images in presentations. Some organizations will find value combining both approaches.

Customized

There are multiple techniques to "tune" or customize a model. Prompt-tuning and reinforcement learning with human feedback (RLHF) are the fastest and most common methods customers are using to train foundation models.

Proprietary

Building a foundation model from scratch requires deep expertise from AI/ML researchers and a deep investment. The vast majority of enterprise organizations will find a higher ROI and faster time to market by customizing existing models across open source and vendor offerings.

Key questions:

Do we have the in-house knowledge, talent, or need to hire accordingly?

Are our current contracts flexible enough to support gen AI exploration / use cases?

Do we have quality, labeled data to customize models effectively?

Do we have governance in place across data and ML models to successfully manage and version control custom models?

Have we done a cost-benefit analysis of customizations and performance?

Google

Right-size your models

Choosing the right model is critical to optimizing the performance, capabilities, and cost of your AI applications. There's no single defining LM or foundation model for all use cases. Some general-purpose models are trained on huge amounts of data that are irrelevant to your needs — for example, your organizational model doesn't need to know the words to every Taylor Swift song to generate a summary report on population health trends or next quarter's sales goals by region.

The compute costs and energy requirements of generative AI models can be high — which is why choosing the right platform and model for the job are imperative. Think carefully about just how much IQ your model actually needs, and be aware that some vendors and solutions are more sustainable than others.

Key questions:

Do we have the capability to evaluate multiple models?

Do we have an AI partner that offers us choice and flexibility with model selection?

Establish an AI governance process

In 2019 Google recognized the need to build out a clearly articulated set of AI principles that would dictate the types of services and benchmarks Google would use for AI in its own businesses.[27]

Today, those principles continue to guide our work including how we map to key frameworks, like the National Institute of Standards and Technology (NIST) AI Risk Management Framework, and support such efforts at the AI Commitments we recently joined with the White House.[28] In this new era of AI transformation it is key that public sector organizations have thought through and determined their own principles and procedures for use of AI in their organization.

When a new technology moves as fast as generative AI does, it can be hard to keep up. As a strategic partner to our customers, Google Cloud helps public sector Leaders chart their path forward with the appropriate frameworks, tools, and governance structures — and ingrain a responsible, consciously cautious approach to AI across your organization.[29]

Developing a responsible AI framework means addressing complex issues of fairness, interpretability, privacy, and safety and security. Within Google, we've built an AI and advanced technologies governance program that is aligned with the NIST approach and underpinned by industry-leading research and a growing library of resources, tools, and recommended practices. As the impact of AI increases across sectors and societies, it is critical to work towards systems that are fair and inclusive for all. As AI advances in making predictions and decisions, interpretability– how AI explains its own logic–becomes a key factor in establishing trust and reliability. As AI analyzes more and more data, it must ensure the privacy of individuals and allow them to control access, whether through legal and regulatory requirements or institutional safeguards. Finally, as AI develops, unexpected scenarios will arise that require new approaches to safety and security.

27. Google AI Principles
28. AI Commitments Factsheet
29. Google Responsible AI Practices

In the face of these challenges, here are some of our recommended best practices:

Does building in-house give us a new unique service? How so?

_____

What option creates impact through simplicity?

_____

What's the fastest way to go live?

_____

Do we have engineering / PM experience & core competencies needed to build it?

_____

What's the cost of building vs. buying?

_____

What's the cost of building vs. buying?

Innovate faster with generative AI for organizations

Google is an AI-first company. Having already built some of the industry's leadingAI capabilities, we continue to focus on making it easy and scalable for everyone to innovate with AI.

We support the needs of generative AI in your organization with:

We have the most comprehensive platform now available and **ready-to-go** with strong support from leading organizations; helping companies create amazing content, synthesize and organize information, automate processes, and build engaging experiences.

You data is *your* data

We do not use our customer's data to train Google's models. The question we hear most is "Do I have control of my data, brand, IP risk, and ability to meet regulatory requirements?"

Everyone can be an AI developer

All users with varying levels of expertise can create innovative enterprise search, chat, and vision apps. We enable both biz & tech practitioners to be more productive using AI assistants.

We deliver infrastructure that is optimized for AI workloads by providing organizations access to the latest GPUs and TPUs, a rich choice of deep learning VMs, and the ability to easily build custom AI software.

The Google Cloud AI portfolio supports all stages of your generative AI journey. With a rapidly growing suite of generative AI technologies being made available — along with new educational and consulting programs, blueprints for specific industry use cases, and our growing partner ecosystem — we are ready to get you and your teams learning, building with, and deploying generative AI.

About Google Cloud

We help organizations innovate faster with enterprise-ready generative AI.

With decades of enterprise AI experience behind us, we are proud to be the AI partner of choice for many leading organizations. Already, many public sector organizations are using our AI platform to deliver on the mission. We look forward to helping you hit the ground running with generative AI, too. To learn more about our suite of generative AI tools, visit **goo.gle/cloud-gen-ai**

About Google Public Sector

Google has been working with government and education since 2006.

That year, we expanded our offerings beyond consumers, launching solutions purpose-built for enterprises. In 2022, we officially launched Google Public Sector, a separate subsidiary that focuses on helping U.S. public sector institutions accelerate their digital transformations to advance their mission.

We help federal civilian agencies, defense and intelligence organizations, state and local governments, and educational institutions serve their publics, modernize their infrastructure, improve operations, enhance cybersecurity, and meet compliance requirements.

We continue to make significant investments and grow our team to meet the increasingly complex needs of government and education.
To learn more about Google Public Sector, visit **goo.gle/about-ps**

About the Google Cloud Generative AI Benchmarking Study

The Google Cloud Customer Intelligence team conducted the Google Cloud Generative AI Benchmarking Study in mid-2023. Participants included IT decision makers, business decision makers, and CXOs from 1,000+ employee organizations considering or using AI. Participants did not know Google was the research sponsor and the identity of participants was not revealed to Google.

Helpful resources

To browse more case studies

Case studies and stories from grade school to grad school →

Government Customer Stories →

To explore courses and training programs

Generative AI on Google Cloud: New training content, from introductory to advanced →

Skills Boost: Introduction to Generative AI Learning Path →

To learn more about our Responsible AI practices and principles

Google Responsible AI Practices →

Google AI Principles →

Google