6 steps to success with generative AI

A practical guide for organizations to make their artificial intelligence vision a reality

Table of contents

INTRODUCTION

Forging ahead 3

What are artificial intelligence, machine learning, and generative AI? 4

Why generative AI? 5

THE GENERATIVE AI JOURNEY

Step 1: Championing an innovative culture 7

Step 2: Make data your differentiator 9

Step 3: Finding the right problem to address 10

Step 4: Upskilling teams 12

Step 5: Scaling beyond pilot projects 15

Step 6: Measuring the results 17

CONCLUSION

Taking the next step with AWS 19

Solving the biggest artificial intelligence challenges 21

INTRODUCTION

Forging ahead

When deployed with the right strategy, generative artificial intelligence (AI) has the potential to change the anatomy of work. By streamlining work activities that absorb 60–70 percent of employee time, organizations can improve productivity down to the individual worker.[1] This can increase agility, streamline processes, boost revenue, and allow for better, faster decision-making.

The lack of a single universally accepted playbook for generative AI success is keeping some organizations on the sidelines, unsure of how to take the next (or even the first) step on the AI journey. By democratizing generative AI, Amazon Web Services (AWS) is helping organizations overcome barriers to adoption and forge ahead with confidence. This eBook outlines a proven path—from taking the first step to measuring results—with insights from AWS best practices and its experience helping thousands of customers realize their own initiatives.

Generative AI could add

$2.6 TRILLION

to

$4.4 TRILLION

in economic benefits annually

The impact of generative AI on productivity could add trillions of dollars in value to the global economy, increasing the effect of all AI by 15%–40%[1]

[1] "The economic potential of generative AI: The next productivity frontier," McKinsey Digital, June 2023

What are artificial intelligence, machine learning, and generative AI?

AI is a way to describe any system that can replicate tasks that previously required human intelligence. Almost all AI systems are built using machine learning (ML). ML utilizes large amounts of data to create and validate decision logic, forming the basis of an AI model. The AI application then feeds input data into that model, and the model outputs humanlike decisions. The rapid advancement of ML, massive proliferation of data, and easy availability of scalable compute capacity can help businesses accelerate how they use AI and, particularly, how they use generative AI.

Generative AI is a type of AI that can create new content and ideas, including conversations, stories, images, videos, and music. Like most AI, generative AI is powered by ML models—very large models that are pretrained on vast amounts of data and commonly referred to as foundation models (FMs).

Why generative AI?

Before diving into the steps of your generative AI journey, let's explore why organizations should begin their journey in the first place. According to Goldman Sachs, generative AI could drive a 7 percent (or almost $7 trillion) increase in global GDP and lift productivity growth by 1.5 percentage points over a 10-year period.[2] Globally, we have reached an inflection point where most user experiences and applications will be reinvented with generative AI.

Developing a strategy that includes generative AI initiatives is imperative to ensure successful organizational outcomes. Even after completing the steps outlined in this eBook, an organization will need to regularly remind itself what it's working toward—staying focused on the precise organizational benefits that can be unlocked by fully harnessing generative AI technology.

[2] "Generative AI could raise global GDP by 7%," Goldman Sachs, April 2023

Organizations are already realizing the impact of:

Boosting employee productivity

Generative AI is being used for its transformative value to help organizations reach new levels of productivity. The latest advancements in generative AI can be used to increase employee productivity with the help of generative AI–powered conversational search, content creation, text summarization, and code generation.

Improving user experiences

Today's public sector organizations can take advantage of generative AI to improve user engagement, increase personalization, and attract new users through deeper experiences. AWS customers have improved their own customer experiences through the effective use of chatbots, virtual assistants, intelligent contact centers, and personalization. In 2023, Spain's public broadcaster, **RTVE**, improved its customer experience using **Amazon SageMaker** to quickly generate necessary content in 2 hours as part of RTVE's coverage of the general election in Spain.

Transforming content creation for greater creativity

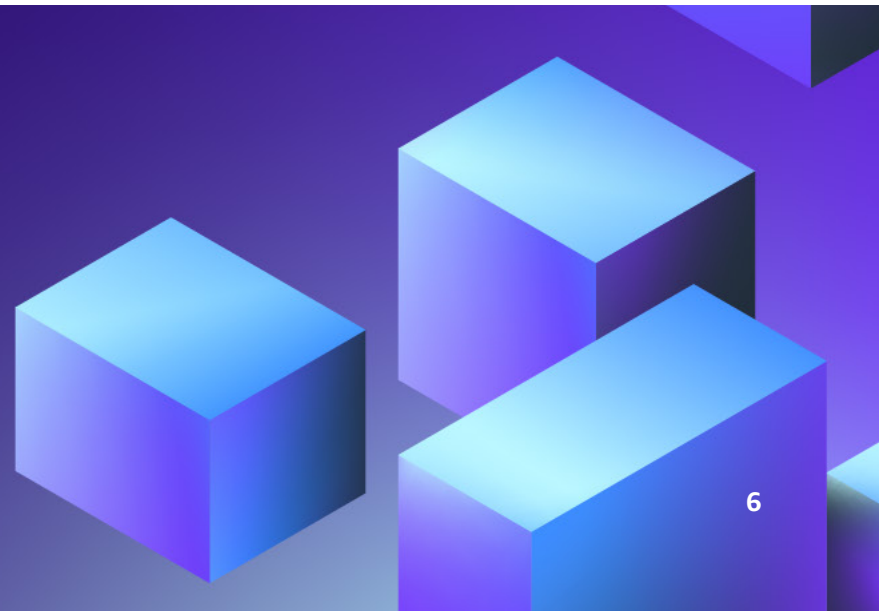
The AI-powered capabilities for automating writing, media design, and character modeling allow for unprecedented creative exploration and speed. Companies like game-industry-enabler **Scenario** are using generative AI to turbocharge real-world production processes across multiple types of creative content—including art and music, text, images, animations, and video.

Driving outcomes for real, competitive advantage

Regardless of an organization's field and domain, generative AI should be embedded into every organization plan to stay relevant and competitive. **LG AI Research** successfully deployed a 300-billion-parameter FM that uses both images and text data.[3] The multipurpose model can carry out a range of tasks, helping to boost LG AI Research's competitive advantage across multiple industries.

Now that we've outlined the *why* of generative AI, it's time to explore the *how*

The next sections will demonstrate the steps in the AI journey using AWS best practices and those of AWS customers to exemplify the necessary changes that must take place in order to successfully implement, deploy, and scale AI solutions.



[3] "LG AI Research Develops Foundation Model Using Amazon SageMaker," AWS Customer Story, 2023

The generative AI journey

The generative AI journey is not necessarily a straightforward path. Achieving success with generative AI requires more than great technology—it also means ensuring the organization is aligned to the right goals. Identifying and reaching those goals will necessitate broad changes in processes, management, and culture. The next sections will explore how organizations can overcome common challenges that often impede progress and take the right steps to implement AI in efficient, sustainable ways.

STEP 1

Championing an innovative culture

Unlocking the full business potential of generative AI requires cultural changes in team organization, objectives, and outlook.

For generative AI to proliferate through an organization, both adminstrative and technical teams must work together and share the same priorities. To achieve this at the outset, the generative AI effort must be supported from the highest levels, with goals set by executive champions and an investment in the technology and processes that enable success. This includes a commitment to build more responsible AI from the beginning—working to identify and mitigate bias, improve explainability, and help keep data private and secure. By taking a people-centric approach, organizations can work to educate their workforce on responsible AI and build more diverse teams to bring more perspectives to the table to improve fairness.

It's important for management to take a wide-scale view while fostering AI initiatives. Leaders must be firm in their goals, but also flexible in how the organization reaches them. Mistakes are sure to be made. But by staying focused on the long-term outlook and minimizing discouragement, organizations can glean wisdom from every error and apply those learnings to champion an innovative culture throughout the business.

Perhaps the largest cultural change organizations must undergo is utilizing the opportunity inherent in mistakes. AI is an iterative process that can only succeed through constant experimentation. Often, these experiments will result in failure. Only by learning from mistakes—and refusing to grind progress to a halt in the name of determining "what went wrong"—can organizations consistently reach the breakthrough successes waiting on the other side.

As new generative AI applications are deployed, organizations should provide training to employees on how to use generative AI responsibly and effectively. Organizations should also have clear policies guiding appropriate usage, while leaders should encourage experimentation with generative AI while managing risks. When implemented thoughtfully, generative AI can augment human capabilities and unlock innovations.

How Amazon did it

Amazon has been using AI for over 20 years. After a decade of experience with the technology, our leadership team asked every business leader in the organization—irrespective of whether they ran a research team, a fulfillment center, or an HR organization—to answer the question of how they planned to use AI in their businesses.

"We don't plan to" wasn't an acceptable answer in most cases, which forced the leadership, domain experts, and technical experts to collaborate on AI initiatives and let nothing halt their progress—even in instances where tangible benefits were still years down the road.

In addition to hiring external data scientists, Amazon created the **Machine Learning University** (MLU), which trained many of its developers to use AI more effectively. The company also built tools like **Amazon SageMaker**, which simplifies model creation and removes the barriers to entry, enabling AI technologies and initiatives to scale more effectively. Additionally, Amazon created a set of pre-built AI services that provides ready-made intelligence to address common business use cases—without customers having to build their own models. For example, **Amazon Bedrock** is a new service that makes generative AI FMs from AI21 Labs, Anthropic, Cohere, Meta, Stability AI, and Amazon accessible via an API. Amazon Bedrock also is the easiest way for customers to build and scale generative AI–based applications using FMs. Amazon Bedrock offers the ability to access a range of powerful FMs for text and images—including **Amazon Titan** FMs—through a scalable, reliable, and secure AWS managed service.

This is why major organizations—including Intuit, Thomson Reuters, AstraZeneca, Ferrari, Bundesliga, 3M, and BMW—as well as thousands of startups and government agencies around the world, are transforming themselves, their industries, and their missions with generative AI solutions from AWS. We take a democratizing approach to generative AI, and we work to take these technologies out of the realm of research and experiments and extend their availability far beyond a handful of startups and large, well-funded tech companies.

Let's take a look at some examples of how Amazon is leveraging AI.

Amazon uses AI throughout its fulfillment process, leveraging a forecast system that can predict demand for nearly every product in its enormous inventory. These prediction models allow Amazon to better deliver on customer expectations of convenience, cost, and delivery speed.

"We forecast millions of products every single day across all of our Amazon sites worldwide," said Jenny Freshwater, Director of Forecasting at Amazon. "And without machine learning, we would not be able to produce those forecasts."

The examples go on and on. Amazon also created Alexa, which provides customers with an entirely new way to interact with technology. Additionally, the company developed groundbreaking technology with autonomous flight via Amazon Prime Air drones. And Amazon also uses robotics in its fulfillment centers to get packages to customers faster.

Achieving these successes required great investments in technology, research, and talent. But those investments would have gone to waste without the cultural changes that pushed them forward through many failures and unexpected challenges. Every organization must follow suit and foster this same fault-tolerant culture of experimentation and innovation before the AI journey can truly begin.

Amazon is using AI to minimize packaging waste, reducing outbound packaging weight by

33% and eliminating

915K TONS

of packaging material worldwide.[4]

[4] "How Amazon is using machine learning to eliminate 915,000 tons of packaging," VentureBeat, September 2020

STEP 2

Make data your differentiator

Like with ML, data is an important piece of generative AI. While general large language models (LLMs) can be used "out of the box" for some use cases, when you want to build generative AI applications that are unique to your business needs, your organization's data is your strategic asset. Data is the difference between a generic generative AI application and one that truly knows the purpose of your organization and who you serve in order to deliver a better, more differentiated experience.

While some organizations will build and train their own LLMs with vast amounts of data, many more will use their organizational data to fine-tune FMs for their unique needs or to add context to prompts through Retrieval Augmented Generation (RAG). For example, agencies and organizations across the public sector can use data from sources such as data lakes, databases, and data warehouses to create chatbots that can streamline service delivery, provide automated responses for routine inquiries, reduce inquiry waiting times, and improve the overall experience on both sides of transactions.

Because of this, success with generative AI requires relevant, high-quality data, which means that you need a strong data strategy in the cloud. According to **McKinsey Digital**, "...companies that have not yet found ways to effectively harmonize and provide ready access to their data will be unable to fine-tune generative AI to unlock more of its potentially transformative uses."

The right data strategy for generative AI includes a comprehensive set of services to store and query data at scale, breaks down silos so you have ready access to all of your data for generative AI applications, and makes sure your data is secured and governed throughout the lifecycle of building generative AI applications.

Discover how a **data foundation built on AWS** gives you a strategic advantage when it comes to generative AI.

Typeset is an EdTech company with a focus on simplifying access and consumption of research papers. They are focused on innovation via AI/ML and have built/fine-tuned several generative AI models for summarizing, paraphrasing, and Q&A against the research paper content. All of these are large language models requiring GPU-based machines for inferencing.

STEP 3

Finding the right problem to address

One mistake organizations often make in their AI journeys is employing discrete data scientists who work in silos to build models as proofs of concepts rather than solve real organizational problems. With no specific problem to solve, IT executives will find it increasingly difficult to demonstrate the value of AI projects and initiatives to their administrative executive counterparts. This can stall or even stop progress on AI initiatives.

Here are some important questions organizations should ask before embarking on an AI journey:

1. Is the project important enough to get attention and adoption?
2. Does it solve a real organizational problem?
3. Do we have the right data to solve the problem?
4. Will the project benefit from AI?
5. Can it eventually be operationalized?

The **AWS AI Use Case Explorer** is a business outcome-centric search and navigation site that enables users to find the right AI use cases, discover relevant customer success stories, and mobilize their teams toward deployment.

In a successful AI journey, organizations create AI teams to address specific organizational problems. This requires including both technical and domain experts within these teams. While the technical experts will take on the brunt of model creation, they need the field knowledge of domain experts to define precise business challenges and identify the data most important to finding a solution.

This approach is also critical to change management. When technical and domain experts collaborate to create models, employees will feel more confident in making decisions based on the algorithm's logic. Together, these teams should also work through how to measure success.

For more on measuring the success of AI initiatives, refer to **Step 6** in this eBook.

Organizations can also take advantage of the **AWS Generative AI Innovation Center** to work backwards from business challenges and go step-by-step through the process of creating AI projects and initiatives.

RTVE uses AWS to deliver quality broadcast services across Spain

RTVE, the public broadcaster in Spain, has been harnessing artificial intelligence (AI) since 2018 to deliver quality services to the millions of people across the country. Its latest project—using Amazon Web Services (AWS) AI tools such as Amazon SageMaker—successfully generated 30,000 pieces of content in 2 hours as part of RTVE's coverage of the 2023 Spanish general election.

By automating the creation of graphics and text, and using conversational AI with voice recognition to synthesize speech, it personalized content for the 4,500 less populated municipalities across Spain, mainly in rural areas. Pere Vila, technology strategy director at RTVE, says the broadcaster is looking forward to greater generative AI innovation to provide quality radio and television services using AWS in the coming years.

Read the full story ›

"In the Election program, we have produced 30,000 news content produced in 2 hours, this content is produced as RTVE produces: leaded by our journalists, with our language, with our values, with our ideas."

Pere Vila, Technology strategy director,

STEP 4

Upskilling teams

In parallel with creating a comprehensive data strategy, today's organizations must focus on arming their teams with the right skills to succeed in the era of generative AI. However, organizations across the public sector are growing increasingly aware of an expanding skills gap—the separation between technologies and the ability of internal IT and administrative specialists to take full advantage of them. It's an issue that should set off alarm bells in light of recent research from the World Economic Forum, which found that more than 75 percent of organizations plan to adopt AI in the next five years.[5]

To help individuals train for the AI jobs of the future, AWS released **on-demand skills training** to support those who want to understand, implement, and begin using generative AI. Amazon has also designed training courses specifically for developers who want to use **Amazon CodeWhisperer**.

While there's no one-size-fits-all solution to the AI skills gap, there are proven methods that can maximize the abilities of existing staff, reducing the need to make large investments in buying or borrowing pretrained expert talent.

[5] "The Future of Jobs Report 2023," World Economic Forum, April 2023

These methods include:

Defining the skills gap: Before closing the skills gap, an organization must identify the precise differences between what it needs or wants its employees to do and what its employees currently can do.

Understanding how skills are mapped: Because AI initiatives are interdisciplinary efforts, an organization should map the skills needed across data scientists, data engineers, analysts, application developers, statisticians, and other subject matter experts in the organization.

Customizing training for specific needs: If an organization has existing training curriculums that could be useful, it should work to tailor those materials to the organization's specific AI needs. Leaders should also investigate pretrained AI services that provide ready-made intelligence for applications and workflows.

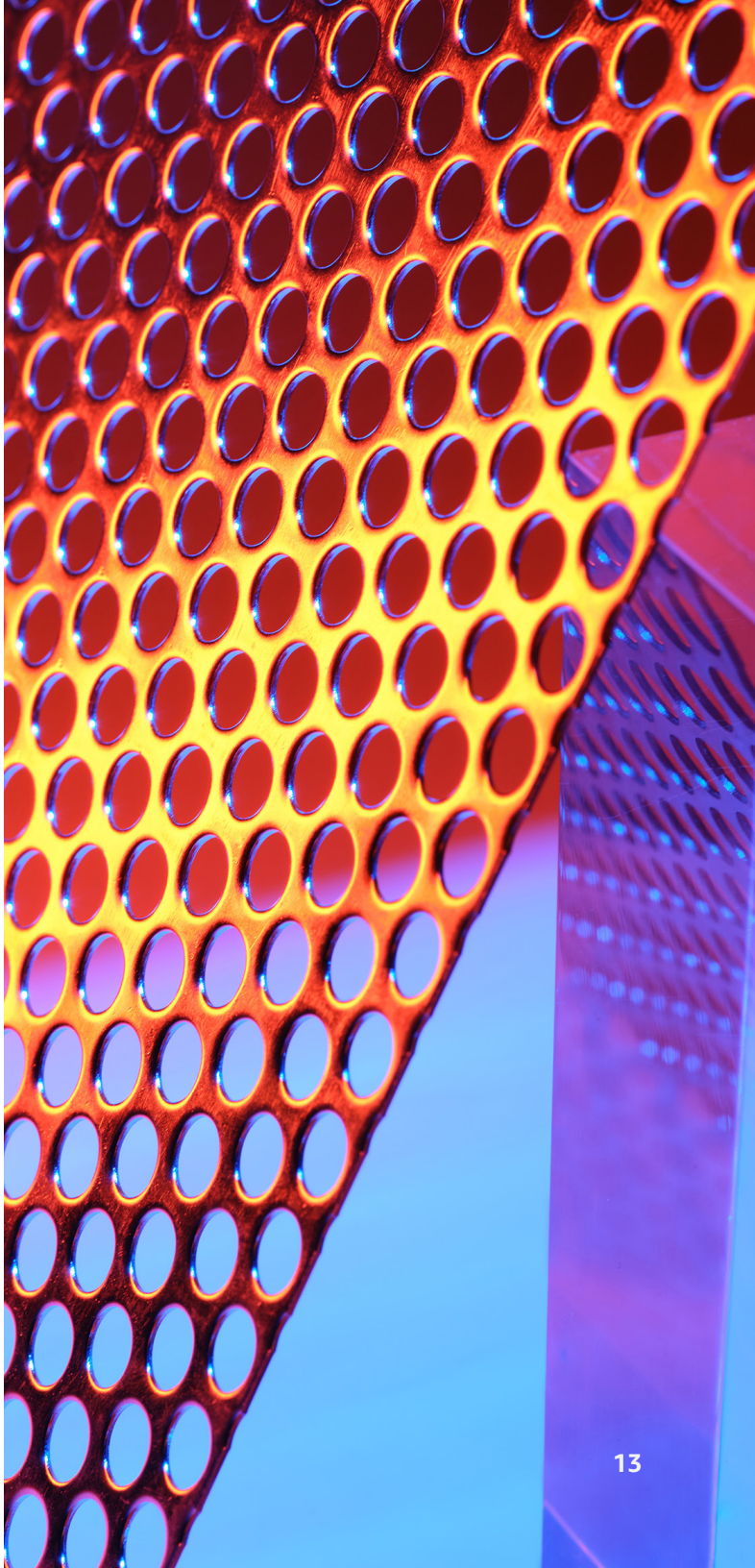
Organizations also need to align teams to successfully tackle AI problems by:

Promoting a culture of empowered teams: AI project teams must be cross-functional, with the authority to execute individual objectives and the freedom to organically cross-pollinate with other teams as demands dictate and opportunities arise. To make this kind of teamwork possible, management will need to embrace new structures—letting go of the strictly hierarchical and departmentally siloed organizational models of the past.

Starting with a pilot team: The organization should establish a pilot team of engineers, IT and AI practitioners, and industry leaders and task it with an AI project.

Enabling organic transformation: Once the pilot project is complete, the organization can split up the team, add people to create new teams, and task them with new projects. This process continues, allowing knowledge to spread organically from veteran team members to new recruits and pollinate between teams.

By following this guidance, many organizations are finding that the people they currently have are actually the people they need to close their AI skills gap. While some recruiting may still be required, investment in the right tools, processes, and management changes can do much of the work to upskill talent for AI success.



How PiecesTech improved clinical workflows

Pieces Technologies will announce that it has fully incorporated the latest generative AI services from AWS. Pieces Technologies is producing “Sculpted AI” using Amazon Bedrock, Amazon SageMaker Canvas, and Amazon Elastic Compute Cloud (Amazon EC2) services to produce “rapid and sculpted application of AI for in-electronic medical record (EMR) clinical workflows” for all its health system and clinical business partners across the US and the world. With AWS generative AI services, Pieces has further developed its state-of-the-art system, the first to autonomously draft over a million inpatient clinical summaries in real-world clinical environments earlier this year.

Pieces uses collaborative AI, auto-prompt engineering, and human in-the-loop systems control to maximize safety and promote rapid iteration in response to customer feedback. Pieces has an industry-leading “severe hallucination rate” approaching one error per million clinical summaries, which it contractually guarantees through its Pieces SafeRead program.

Read the full story ›

STEP 5

Scaling beyond pilot projects

After the first few successful pilots, organizations must take the next step on the journey: sustainably scaling AI across the business. This is both a technical and a cultural challenge. There are several ways companies approach the cultural shift necessary to scale AI. Some might find success by creating a center of excellence that rallies the community and continues to push for new initiatives. Or, like Amazon, organizations can make AI an integral part of yearly planning processes, continuously bringing domain and technical experts together to brainstorm and determine their next steps.

Achieving scalability requires organizations to help their developers use AI. Building models at scale can be labor intensive, which can slow innovation. With **Amazon Bedrock** serverless experience, you can get started quickly to privately customize FMs with your own data and easily integrate and deploy them into your applications using AWS tools and capabilities you already know (including integrations with features like pipelines to manage your FMs at scale). Many organizations are solving scalability with SageMaker to prepare data and build, train, and deploy models to get them into production faster and at lower cost, enabling sustainable expansion of AI initiatives beyond pilot projects.

Many more organizations are scaling through **AWS AI services**, a set of pretrained and managed services that can be used as building blocks to address common use cases, including personalizing recommendations, modernizing their contact centers, improving safety and security, and increasing engagement with those they serve.

How the Allen Institute did it

Using AWS generative AI tools, the **Allen Institute** is mapping the entire human brain. In the process, its researchers are creating the largest open source database in the world. Named the Brain Knowledge Platform, the project is helping to accelerate progress in treating brain diseases such as Alzheimer's and Parkinson's.

"We need this map to know exactly what goes wrong in disease and precisely where to target therapeutics," says Dr. Ed Lein, Senior Investigator for the Allen Institute. "Also, the generative AI side of things has some unusual possibilities for us. We'll be able to integrate data of many different types to begin to make inferences that human researchers are not capable of doing."

"I think we can imagine the new territory that we're entering with generative AI, as it may be applied to neuroscience. We can begin to understand how the brain actually works—and to harness this knowledge for the good of society."

Dr. Ed Lein, Senior Investigator, Allen Institute for Brain Science

Watch the video ›

STEP 6

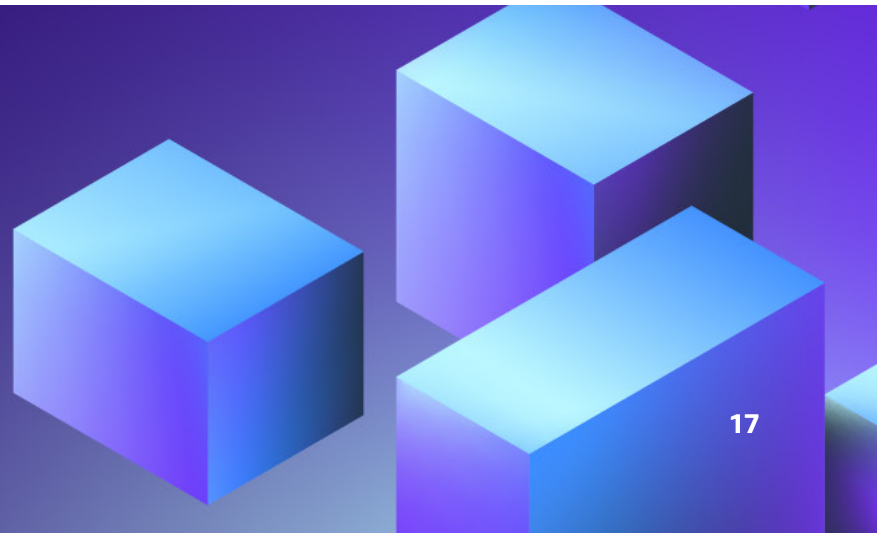
Measuring the results

When measuring the results of AI efforts, the traditional "project ROI" viewpoint—where a project has defined start and end points, a budget, and a return—is reductive and can be detrimental to the initiative's success. If the project doesn't generate a positive return within the given time frame, the organization may lose interest and miss out on critical opportunities down the line. Instead, executives and IT alike must measure AI efforts based on what success means for their organization with regard to the processes being optimized. In addition, they must view AI efforts as long-term investments, acknowledging that a true "return" may not be realized for several years and after countless iterations.

When planning AI initiatives, it's better to view the process through the lenses of agility, competitive advantage, or risk tolerance rather than "expected" return. An organization will have greater success if it disregards the question of "What will be my value in X months?" in favor of something more like, "If we don't invest in this now, how far will we fall behind in X years when the technology matures?"

While the impact of AI initiatives can still be measured—it just requires a different outlook. AI results can be measured through something resembling a "value tree," where the main trunk of the tree represents the traditional revenue return and branches extending from the trunk recognize the value of other organizational outcomes. The specific branches of the value tree will depend on the organization, the industry, and the initiative. But they might be things like time saved through automated processes, constituent value-creation, customer service improvements, or application development acceleration.

Measuring the success of AI through a more holistic and long-term model will keep organizations focused on the best outcomes for their future.

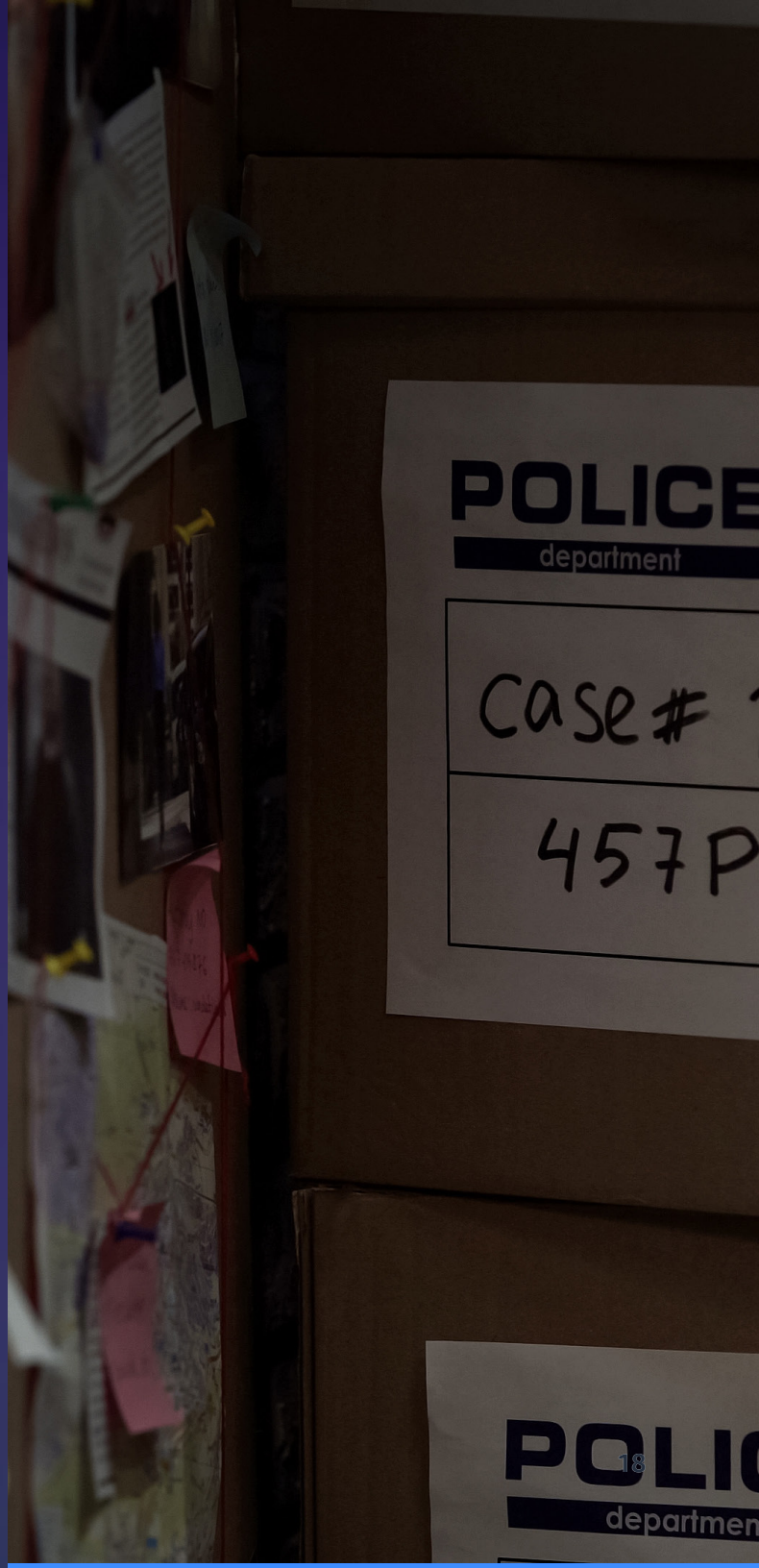


Bedfordshire uses generative AI to redact case files

In 2021, **Bedfordshire Police** signed the first of its kind collaboration between a law enforcement agency and AWS to address the high-priority issue of redaction of case files. The inefficiencies within the current process led to a significant waste of police officer time and municipal budget.

In 2023, AWS and UK software company Riven began working with Bedfordshire Police to examine automated redaction of documents and how they could harness AWS services to quickly solve the issue. The result, Riven DocDefender, was far superior to any other on the market, saving time equivalent to at least 5 full-time police officers.

Through the platform's quick integration and zero requirement for training, Bedfordshire Police had fully adopted the software across the entire organization within 16 weeks of commencing the work. Since then, other UK police forces have followed and shared in the same success.



CONCLUSION

Taking the next step with AWS

No matter where your organization is in its AI adoption journey, you can take the next step with AWS solutions built on the most comprehensive cloud platform and optimized for generative AI with high performance computing (HPC), security, and analytics. Featuring the world's broadest and deepest set of AI services, over 100,000 customers are running their AI workloads on AWS. Generative AI can help you realize new business value within your organization. From reinventing customer experiences to enhancing productivity and accelerating growth, generative AI holds the power to help you transform your business.

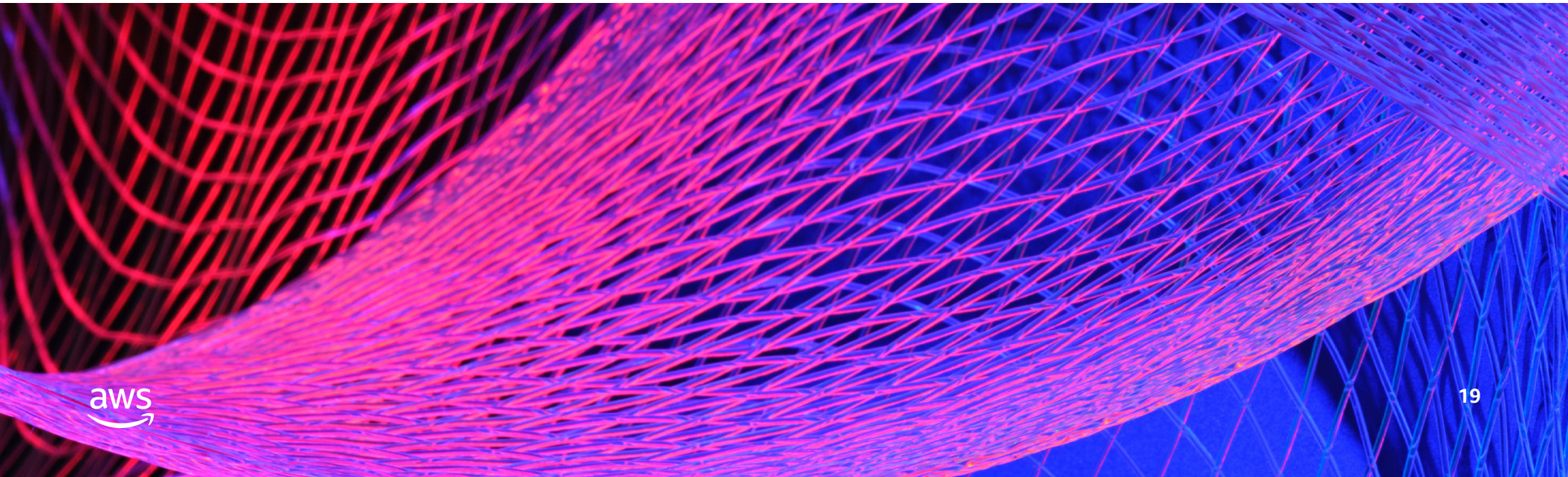
Generative AI with AWS, by the numbers

100,000+

customers using AWS for their AI workloads

20+ YEARS

of building experience at Amazon



Improve customer experiences

Chatbots and virtual assistants: Streamline customer self-service processes and reduce operational costs with generative AI–powered chatbots, voice bots, and virtual assistants to automate responses to customer queries.

Agent assist and conversational analytics: Enhance agent performance to improve first contact resolution and augment tasks such as knowledge search, call summarization, and problem-solving. Managers can extract valuable insights to improve customer experience, monitor agent performance, and boost business performance.

Personalization: Deliver better-personalized experiences and increase customer engagement with individually curated offerings and communications.

Boost employee productivity

Conversational search: Improve employee productivity by quickly and easily finding accurate information and summarizing content through a conversational interface.

Code generation: Accelerate application development with code suggestions based on the developer's comments and code.

Automated report generation: Generate reports, summaries, and projections automatically, saving time and reducing errors.

Enhance community engagement

Broaden accessibility: Effectively connect with all members of the community using tools like language translation and interactive chatbots.

Improve information accuracy: Minimize the human impact of inaccurate information and reduce the risk of misdiagnosis by using generative AI to support and enable human reviewers in decision-making.

Enhance creativity in problem-solving: se generative AI to produce a broad range of solutions to complex societal problems and provide diverse perspectives for public officials.

Accelerate process optimization

Document processing: Improve business operations by automatically extracting and summarizing data from documents and insights through generative AI–powered question and answering.

Data augmentation: Generate synthetic data to train ML models when the original dataset is small, imbalanced, or sensitive.

Supply chain optimization: Improve logistics and reduce costs by evaluating and optimizing different supply chain scenarios.

CONCLUSION

Solving the biggest artificial intelligence challenges

Most organizations have made investments and progress in their AI journeys and are exploring the possibilities of generative AI. But many find themselves hitting obstacles along the way, worried that costs and complexities will grow too high as they progress.

Throughout this eBook, we explored the steps to forge ahead and realize the full power of generative AI. To recap, let's look at the biggest challenges we identified along the way, along with a brief recommendation of how your organization can solve them.

To learn more about how you can overcome obstacles and accelerate your AI journey, visit the AWS AI Resource Hub.

To learn more about how generative AI can boost productivity, build differentiated experiences, and innovate faster for every businesses, visit the AWS Generative AI Homepage.

Get started ›

Challenge	Solution
Discouraging failures	Developing a fault-tolerant culture
Siloed, unprocessed data	Creating a modern data strategy that includes data lakes
Finding the right business problems	Building blended teams that include both technical and domain experts
The AI skills gap	Adopting new organizational models, processes, and team management philosophies
Sustainably scaling beyond pilot projects	Taking advantage of end-to-end tools like Amazon Bedrock and SageMaker to build and scale generative AI applications
Measuring the results	Forgoing traditional ROI metrics in favor of agility, competitive advantage, and risk tolerance using the value tree model