



**Deloitte.**

# Delivering results with generative AI

Accelerate productivity and innovation across industries with generative AI solutions from Deloitte and AWS.





## Introduction

Generative artificial intelligence (AI) represents transformative potential for businesses in improving customer service, boosting employee productivity, speeding business operations, and producing all types of creative content.

Where “traditional” machine learning (ML) is about prediction, generative AI is about creation—mimicking human intelligence, contextual awareness, and decision-making. Adding this capability to enterprise workflows can improve how businesses serve their customers.

However, although it doesn’t take tech savvy to use generative AI, specialized tools are required to customize, implement, and integrate it into business operations with data quality and management, sound governance, and organizational alignment built in.

Deloitte and Amazon Web Services (AWS) can deliver these capabilities, while helping reduce costs, simplifying execution, accelerating innovation, and transforming customer engagement in current and new markets. Working with Deloitte and AWS, businesses are bringing generative AI out of the experimentation phase and into the real world of delivery—with results.

**Where “traditional” machine learning is about prediction, generative AI is about creation—mimicking human intelligence, contextual awareness, and decision-making.**

## Table of contents

### See generative AI at work

The stories in this eBook are a selection of examples of how Deloitte and AWS deliver results with generative AI solutions across industries, including but not limited to:

Telecommunications	04
Biotechnology	08
Banking and financial services	12
Insurance	16
Wealth management	20



# Centralizing internal documentation with generative AI helps an internet company speed innovation

A company that builds and maintains broadband networks understands the benefit of picking up the pace. Serving over 1.5 million homes and businesses, Chorus New Zealand has expanded the country's fiber broadband network, bringing high-speed internet access to both major cities and non-urban communities.

---



## Streamlining internal information with generative AI empowers software architects to bring customers improved services faster.



To support continued growth and reliable service, Chorus sought to improve efficiency internally. Specifically, it wanted to leverage generative artificial intelligence (AI) to enable its software architects to access and understand internal information more quickly, improving productivity, speeding decision-making processes, and accelerating time to market for new software products they wanted to build. For example, adding a new feature to the customer-facing mobile app or building applications to help the company determine optimal paths for laying new cables.

Previously, company documentation was difficult to accumulate, read, and analyze. This hindered the company's ability to escalate and address customer complaints.

Chorus selected Deloitte to build a generative AI solution on Amazon Bedrock that could aggregate software architecture documentation across multiple sources and formats into a unified, searchable database. This enabled software architects to access required information

in seconds. So, instead of spending hours reading and interpreting source code, a software architect could enter a quick text query, such as "Explain what this software component does" and immediately receive a natural language response, including links to additional source materials. From there, the user can validate the information, ask follow-up questions, and rate the quality of the first generated response. The solution saves feedback in order to improve performance in the future. The tuned generative models can operate with high accuracy once established, requiring only periodic reviews.

Although initial investment was required for developing the solution's capabilities, costs were surpassed by efficiency gains, including a 75 percent boost in architect productivity.

Aggregating architectural documentation with generative AI delivered faster access to information, improved productivity and efficiency, and boosted speed of decision making, helping Chorus continue to connect New Zealanders to the world via high-speed, reliable internet access.

## Project highlights

### Bias for Action

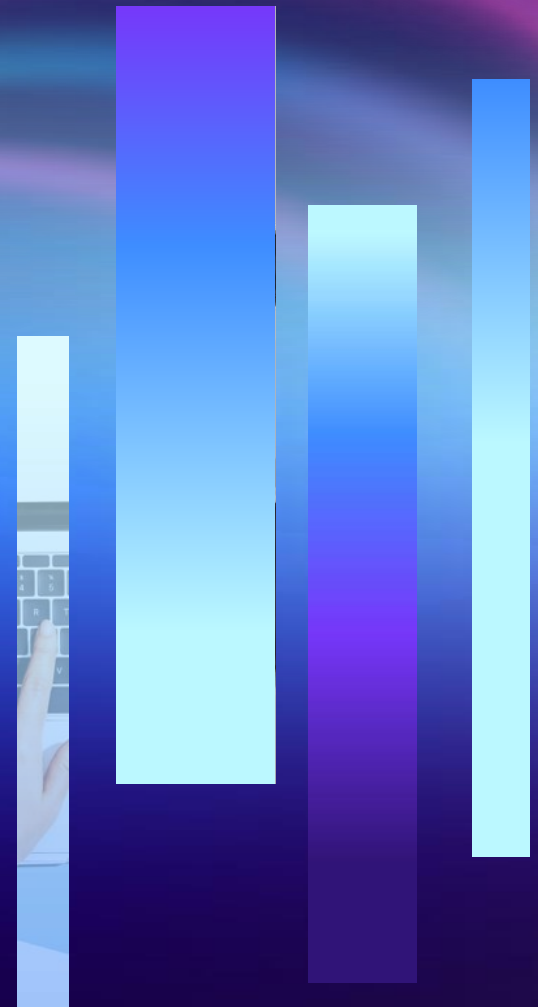
Chorus New Zealand wanted to leverage generative artificial intelligence (AI) to enable software architects to access and understand internal information more quickly, so Deloitte built a generative AI solution on Amazon Bedrock to unify and operationalize necessary documents.

---

### Deliver Results

The Deloitte solution aggregated the company's software architecture documentation into one searchable database, delivering faster access to information. This improved architect productivity by 75 percent, which ultimately supports improved service to the company's growing customer base.

---





Generative AI-assisted aggregation and insight generation led to a 75% boost in software architect productivity.



## Built with

Amazon Bedrock

Amazon Kendra

Amazon Lex

Amazon Simple Storage Service (Amazon S3)

AWS Lambda

# Advancing innovation in biotechnology with a progressive generative AI trial

RNA technology supports replacement therapy, vaccines, and cell therapy—all critical tools in fighting serious diseases<sup>1</sup>. Alnylam, a leading biotechnology firm working in this field is developing a robust pipeline of investigational medicines that could help treat patients living with both rare and common diseases of the eye, liver, cardiovascular system, or central nervous system.

<sup>1</sup> Beyond COVID-19: Unlocking the power of mRNA vaccines and therapies - Thoughts from the Centre | Deloitte UK





## Embracing generative AI to explore investigational medicine more efficiently.

Alnylam selected Deloitte to build out a comprehensive technology strategy for generative artificial intelligence (AI) to help aid and accelerate its work.

Initially, Deloitte created an application for analyzing data collected by physicians or pharmacists who receive customer feedback and complaints from patients about side effects, efficacy, or interactions with other drugs. One challenge in this work is that feedback comes in varied forms, including handwritten notes. Able to detect patterns in patient reports across different formats, the application enabled faster issue resolution and proactive product improvements, such as reassessing the components of a drug to see if side effects can be brought under control—or whether a drug should be taken off market.

Deloitte also created a digital assistant application for employees, supported by a proof of concept developed by the [AWS Generative AI Innovation Center](#). The digital assistant, powered by Anthropic Claude foundation model on Amazon Bedrock, provided quick access to large

volumes of internal documentation, improving employee productivity and reducing inquiry tickets to IT and HR.

After establishing its first two use cases for generative AI, the company onboarded domain experts, facilitated use case workshops, and galvanized stakeholders enterprise-wide to identify higher value automation possibilities.

Working with Deloitte helped Alnylam to establish frameworks and resources to sustain AI innovation. Ultimately, the processes improved with AI tools can enable faster times to market for new drugs or improvements, speedier issue resolutions for customer feedback, and overall greater support for patient and drug safety.

## Business outcomes

### Deliver Results

Deloitte developed a generative artificial intelligence (AI) application that helped improve and accelerate analysis of patient complaints about drugs in market.

---

### Insist on the Highest Standards

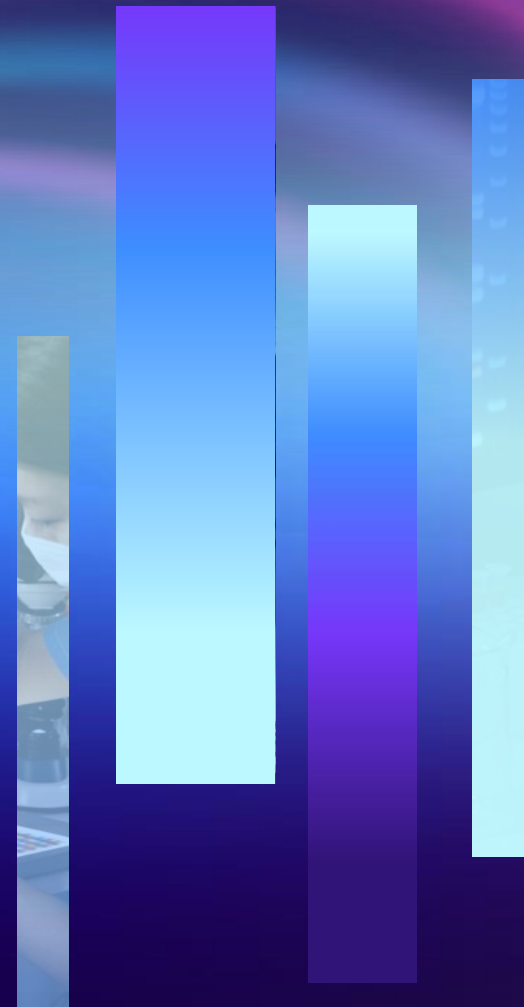
Deloitte built an end-to-end architecture using Amazon Web Services (AWS) to realize the two initial use cases and set up a foundation for future generative AI use cases.

---

### Think Big

Deloitte worked with the Alnylam to stand up a Generative AI Center of Excellence to centralize talent, tools, and leading practices. This drove efficiency and accelerated AI development through expertise sharing.

---





Able to detect patterns in patient reports, the application enabled faster issue resolution and proactive product improvements.

## Built with

Amazon API Gateway

Amazon Athena

Amazon Bedrock

Amazon DynamoDB

Amazon Kendra

Amazon Simple Storage Service (Amazon S3)

Amazon SQS

AWS Generative AI Innovation Center

AWS Glue


AWS Lambda

# Translating a stream of customer comments into a strategic to-do list—at scale

A Canadian banking and financial services company needed help to understand and act on customer feedback at scale. The company receives reviews and comments across several channels, including app stores and customer satisfaction survey scores (CSATs). However, while serving roughly 13 million North American customers the volume of reviews can quickly grow, making it difficult to discern what warrants a priority response.

---





## A bank leverages generative AI to identify the most urgent customer feedback in real time.

Manually categorizing customer complaints previously took time and limited insights. For example, if dozens of customers used different channels to report errors uploading credit card or loan application documentation through the mobile app, the issue may not become immediately apparent to the bank's document processing team.

So, the bank selected Deloitte to develop a generative artificial intelligence (AI) solution that could transform the customer insights process, driving efficiency, enhancing analysis, and empowering internal teams with comprehensive, real-time understanding of the voice of the customer.

The solution automatically ingests customer reviews from across different app stores and converts the reviews into embeddings, or numerical representations that make it possible for the machine learning algorithm to efficiently understand and associate similar comments into trends or buckets. Once the solution sorts, or buckets, issues based on topic and priority, the most important topics with the highest number of complaints are automatically flagged to be addressed first.

The solution can also summarize volumes of text into concise points that can help customer support teams respond faster. And over time, with techniques like prompt engineering and adjusting operation levers of how the review data is ingested and categorized, the quality of results can continually improve. Deloitte also created a web interface where analysts can query and review the most important topics—with the highest number of complaints—automatically flagged.

By automating categorization and enhancing analysis, this innovation powered by Deloitte and Amazon Web Services (AWS) can drive tremendous efficiency gains, including up to 40 percent cost savings. It also gives customers a better experience on a bank's mobile app, so they don't have to spend valuable time on technical interruptions.

## Business outcomes

### Bias for Action

With a solution powered by Deloitte and Amazon Web Services (AWS), a banking and financial services company can process up to five times the amount of customer feedback with minimized human effort. This helps identify and resolve complaints faster and results in an overall better customer experience across the bank's digital channels.

---

### Insist on the Highest Standards

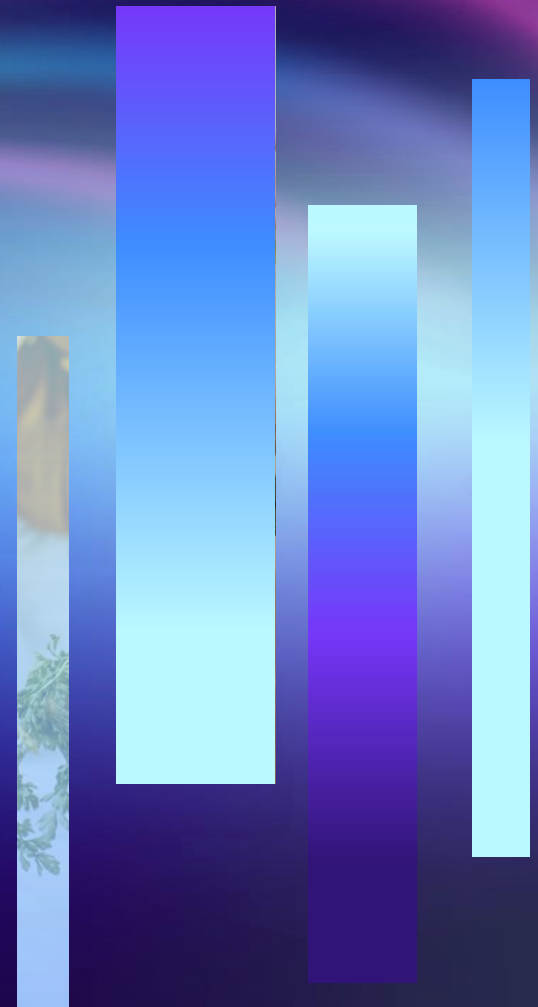
Advanced analytics with nuanced classification across approximately 200 categories and precise sentiment detection helps the company identify trends, such as technical bugs or accessibility issues that need to be addressed immediately.

---

### Deliver Results

The solution has the potential to drive increased efficiency gains, including up to 40 percent cost savings.

---





Generative AI powered by Deloitte and AWS processed up to 5x more customer feedback data with minimized human effort.



## Built with

Amazon Aurora  
Amazon Bedrock  
Amazon CloudFront


Amazon Simple Storage Service (Amazon S3)  
Amazon Titan Embeddings  
AWS Lambda

## Answering customer queries quickly and accurately in the wake of a storm

An insurance company has provided home, car, and business coverage across New Zealand and the Pacific Islands for over a century, with a current base of 320,000 customers. But significant weather events have recently highlighted gaps in the company's ability to provide seamless customer service for claim inquiries during surges in demand.

---





**With help from generative AI, customers can get answers to their urgent policy questions faster.**

In the wake of a storm, customers naturally want timely updates on claim status, repair estimates, and answers to any other questions so they can rebuild as quickly as possible. The insurer receives an estimated 40 percent of such customer requests via email, straining contact center resources.

In order to answer customer policy questions, the insurance company's agents reference an internal knowledge base comprising more than 1,000 pages. Finding the information to answer each question accurately and completely takes time—but after a natural disaster, customers don't have time to wait. Training new agents requires a weeks-long onboarding process, so flooding the call center with additional staff doesn't necessarily help customers with the most immediate needs.

The company selected Deloitte to investigate how generative artificial intelligence (AI) solutions could foster greater contact center efficiency and scale customer service during unforeseen events, without risking accuracy or customer satisfaction.

Deloitte built a proof of concept (POC) to streamline agent responses to customer emails categorized as general inquiries. For example, a customer might email the insurance company to ask how long it takes for purchased flood insurance to take effect. The POC solution can automatically query the relevant knowledge base and create a first draft of the email reply, including citations for quality assurance. Then the agent can quickly fact-check and correct the draft if necessary.

Overall, the generative AI solution successfully provided faster access to key information: customer satisfaction (NPS) scores for email support improved from -15 to +15. It also freed up more time for agents to take on value-added work and even cross-training to serve customers better in the future. Most importantly, customers received the information they needed when they needed it most.

## Project highlights

### Bias for Action

Deloitte built an artificial intelligence (AI) system that can scale automatically to help the insurance company handle surges in email volume following severe weather events, ensuring more timely customer assistance. Agents saved significant time not having to search knowledge bases manually.

---

### Insist on the Highest Standards

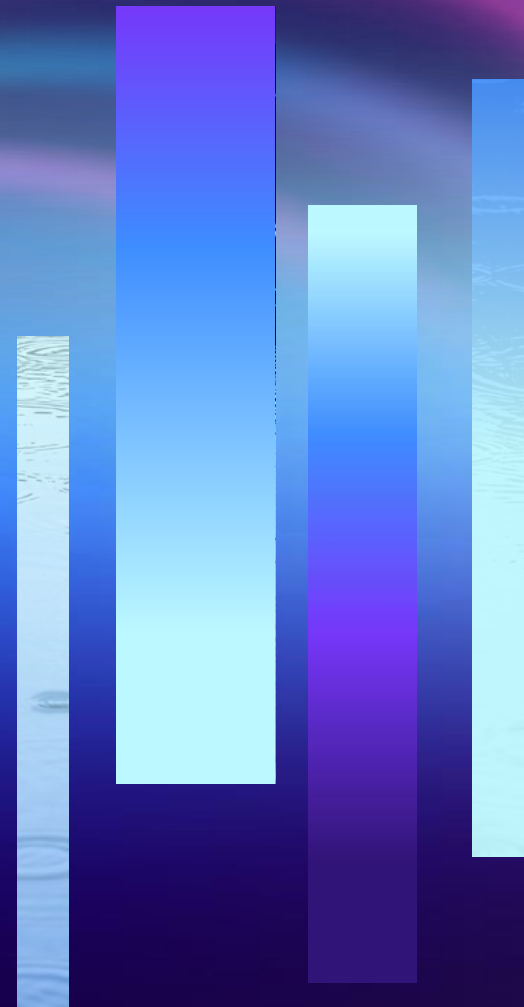
The AI leveraged an internal knowledge database to rapidly research and cite credible sources in response drafts, increasing response accuracy.

---

### Deliver Results

Email response times were reduced by more than 25 percent through AI-generated draft replies. Customer satisfaction (NPS) scores for email support improved from -15 to +15 and 75 percent of users reported an increase in work satisfaction.

---





Finding the information to answer each question accurately and completely takes time—but after a natural disaster, customers don't have time to wait.



## Built with

Amazon Bedrock

Amazon Kendra

Amazon Simple Storage Service (Amazon S3)

AWS Lambda



## Fine-tuning operations to bring customers sound financial advice

Financial markets move quickly, and financial advisors must match the pace. One of the largest investment firms in the United States provides retirement planning and wealth management services to employees of educational institutions, businesses, and other organizations around the world. Managing millions of accounts and more than a trillion dollars, the firm needed an innovative way to scale customer service efforts without sacrificing quality or customer trust.

---



## Leveraging generative AI to boost employee productivity and strengthen customer service.



The firm selected Deloitte to explore how generative artificial intelligence (AI) could help improve the investment strategy and advice delivered to customers. Deloitte developed an application to help the firm's consultants streamline their research as they sought to glean insights from client financial documents.

Because the solution leverages large language models (LLMs) via Amazon Bedrock and machine learning with Amazon SageMaker JumpStart, consultants can pose natural language questions and receive answers almost instantly. For example, if a customer wants to understand what funds they should invest in or why their return is lower, consultants can easily query the generative AI solution and provide informed answers faster. It also reduced the need for firm consultants to switch between systems to find information, increasing productivity. Moreover, the Deloitte solution enabled new use cases, such as asking complex questions and receiving answers and insights drawn from multiple documents.

Behind the scenes, the unified platform consolidated previously separate back-office operations into a single environment, reducing IT overhead for managing integrations and duplicate data. An additional benefit was faster onboarding of new employees who previously struggled with navigating siloed systems. By centralizing infrastructure, scaling to support more transactions and services was also simplified.

Overall, the standardized interface and consolidated backend led to higher employee productivity and satisfaction, positioning the firm to better support customers through all the market's movements.

## Project highlights

### Bias for Action

To help a financial management firm deliver more responsive customer service, Deloitte developed an internal application leveraging generative artificial intelligence (AI) that enables employees to pose natural language questions about customer records and glean insights faster.

---

### Deliver Results

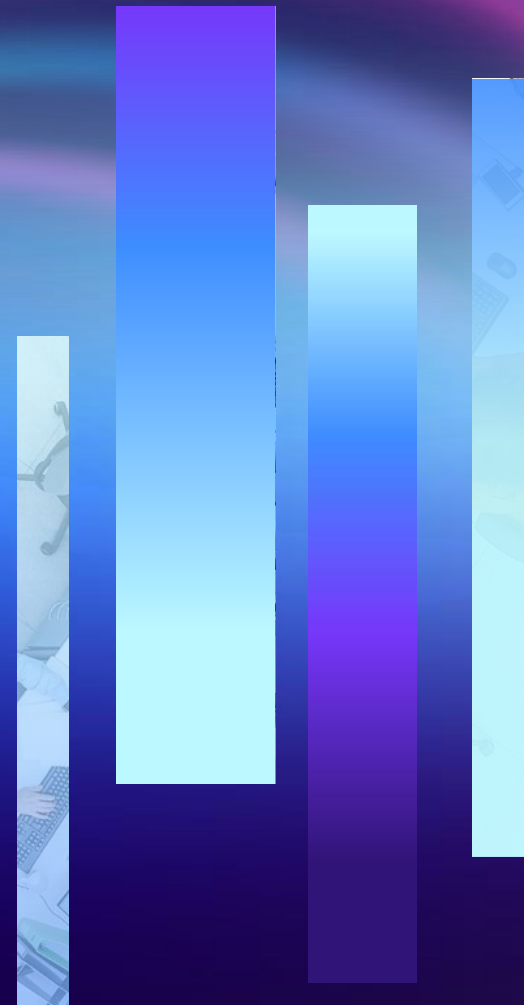
Deloitte built the application with Amazon Web Services (AWS) products that help keep costs and latency low. The modular nature of the application enabled fast iterations on new features, so the firm could continuously enhance how its consultants access and gain insights from the application.

---

### Dive Deep

Robust user security ensured only authorized access to sensitive materials, keeping customer data safe.

---





If a customer wants to understand what funds they can invest in or why their return is lower, consultants can use the generative AI application to provide informed answers faster.

## Built with

Amazon Bedrock

Amazon Cognito

Amazon EKS

Amazon OpenSearch

Amazon SageMaker Jumpstart

Amazon Simple Storage Service (Amazon S3)

Amazon SQS

Amazon Textract

AWS DynamoDB

# Generating innovation with AWS

With AWS, you don't need to invent nor reinvent generative AI capabilities—they are ready to deploy now. You just have to decide what business capabilities to reimagine first.

## START HERE:

Generative AI projects can benefit from these powerhouse tools.



### Amazon Bedrock

The easiest way to build and scale generative AI applications with foundation models.



### Amazon Q

An AI-powered assistant that reimagines experiences across the entire development lifecycle.



### Amazon SageMaker JumpStart

Deploy prebuilt machine learning solutions in just a few clicks.

Learn more about [generative AI on AWS](#).

As depicted in these applications, please reach out to learn more about the possibilities of generative AI for business.

**Kuldeep Singh**  
Principal Global PDM AI/ML,  
Analytics, Database, AWS  
[ksingh@amazon.com](mailto:ksingh@amazon.com)  
Mobile: +1 925 699 2341

**Kasi Muthu**  
WW Data and AI/ML  
Solutions Architect, AWS  
[kasimuth@amazon.com](mailto:kasimuth@amazon.com)  
Mobile: +1 214 533 6159

**Chris Jangareddy**  
Senior Managing Director, AI,  
Data & GenAI, Deloitte  
Consulting LLP  
[cjangareddy@deloitte.com](mailto:cjangareddy@deloitte.com)  
Mobile: +1 214 789 0135

**J.B. McGinnis**  
Principal, Cloud Engineering,  
Deloitte Consulting LLP  
[jbmcginnis@deloitte.com](mailto:jbmcginnis@deloitte.com)  
Mobile: +1 212 436 4615



#### ABOUT AWS

Since 2006, Amazon Web Services has been the world's most comprehensive and broadly adopted cloud. AWS has been continually expanding its services to support virtually any workload, and it now has more than 240 fully featured services for compute, storage, databases, networking, analytics, machine learning and artificial intelligence (AI), Internet of Things (IoT), mobile, security, hybrid, media, and application development, deployment, and management from 105 Availability Zones within 33 geographic regions, with announced plans for 18 more Availability Zones and six more AWS Regions in Malaysia, Mexico, New Zealand, the Kingdom of Saudi Arabia, Thailand, and the AWS European Sovereign Cloud. Millions of customers—including the fastest-growing startups, largest enterprises, and leading government agencies—trust AWS to power their infrastructure, become more agile, and lower costs. To learn more about AWS, visit [aws.amazon.com](https://aws.amazon.com).



#### ABOUT DELOITTE

Deloitte provides industry-leading audit, consulting, tax, and advisory services to many of the world's most admired brands, including nearly 90 percent of the Fortune 500® and more than 8,500 U.S.-based private companies. At Deloitte, we strive to live our purpose of making an [impact that matters](#) by creating trust and confidence in a more equitable society. We leverage our unique blend of business acumen, command of technology, and strategic technology alliances to advise our clients across industries as they [build their future](#). Deloitte is proud to be part of the largest global professional services network serving our clients in the markets that are most important to them. Bringing more than 175 years of service, our network of member firms spans more than 150 countries and territories. Learn how Deloitte's approximately 457,000 people worldwide connect for impact at [www.deloitte.com](https://www.deloitte.com).

"Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see [www.deloitte.com/us/about](https://www.deloitte.com/us/about) for a detailed description of our legal structure.

Certain services may not be available to attest clients under the rules and regulations of public accounting.

This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor. Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

Copyright © 2024 Deloitte Development LLC. All rights reserved.