




Set your document data free

Derive critical insights from your documents—quickly, easily, and at scale—with AWS machine learning services





Many of today's organizations have a large number of documents that are critical to business processes. Extracting, processing, and analyzing data from these documents is traditionally performed through manual systems and tasks—which are outdated, slow, time- and cost-intensive, and difficult to scale. This problem is prevalent across many industries:

Financial services organizations must examine documents, such as loan or credit applications, with the highest degree of accuracy and care, often turning to manual review to pull out sensitive or critical information, such as mortgage rates or credit scores.

Healthcare and life sciences organizations are fighting an uphill battle against an ever-growing mountain of documents and forms, searching for and analyzing data that is essential to clinical trial research and patient diagnosis in order to more accurately treat their patients.

Public sector entities are forced to tie up their already-strained resources to process data from documents, such as applications for a wide variety of services.

Intelligent document processing (IDP) is changing the equation. Powered by machine learning, it can help organizations extract text from millions of documents, understand the sentiment of or relationships between that data, and even include a human step to validate, correct, or augment the machine learning results for higher accuracy and compliance. IDP solutions from AWS can help to empower the employees, clients, and customers of these businesses with fast, easy access to the document data they need—while helping them discover new insights from data that was previously difficult to extract and analyze.

In this eBook, we'll take a deeper look at how IDP can provide practical business benefits—and show you how services from AWS can help.

The benefits of intelligent document processing

IDP is a set of machine learning (ML) solutions that can automatically extract, process, and analyze the document data you need to drive your business. These solutions can play a vital role in helping to overcome outdated, manual, and slow document processing infrastructure.

Organizations across all industries are turning to IDP to unlock a number of key benefits. First, by curtailing or even eliminating errors associated with manual data entry and processing—while still allowing for human intervention when required—IDP helps to **improve data accuracy**.

IDP also enables **faster document data processing**, helping your organization accomplish what once took months or weeks in a matter of days.

IDP can lead to **higher customer satisfaction**, providing your clients with more accurate information and helping you respond to requests faster and more efficiently.

IDP **boosts employee productivity**, helping workers spend more time on business-critical tasks and less time pulling out insights from documents and performing manual data entry.

IDP can also **generate cost savings**. By automating document workflows, it reduces the complexity of data extraction and analysis, allowing your organization to dedicate less of its budget and fewer of its resources to these previously intensive tasks.

In the next section, we'll take a look at how specific AWS services can help you leave the challenges of manual document data processes behind—allowing your organization to achieve these benefits.

Get the most from IDP with AWS

IDP solutions that use AWS machine learning can help you extract text from millions of documents, understand the sentiment of or relationships between that data, and potentially include a human review to validate, correct, or augment output for higher accuracy and compliance.

AWS offers several flexible approaches that can be used to implement a machine learning-based IDP solution, depending on the stage of your document processing journey, business landscape, and desired business outcomes. For organizations that are looking to get started with an IDP solution today, with no ML expertise required, AWS offers services, such as **Amazon Textract**, **Amazon Comprehend**, **Amazon Augmented AI (Amazon A2I)**, and **Amazon Kendra**. Combined together or used separately, these AWS services can provide a powerful way to reduce cost and manual effort—and improve your business outcomes.

Alternatively, organizations that want to develop their own machine learning models for text extraction and analysis can use **Amazon SageMaker**, a fully managed service that helps data scientists and ML developers build, train, and deploy machine learning models quickly.

In the next few sections, we'll review the benefits and general use cases for these services. Later, we'll show how financial services and healthcare organizations, as well as contact centers, are using one or more AWS services for IDP.

Easily extract the data you need— from virtually any document

Organizations looking for a fully managed service to help with accurately and quickly extracting data from documents can turn to **Amazon Textract**. This service uses machine learning to automatically extract handwriting, printed text, and data from scanned documents. It goes beyond traditional optical character recognition (OCR) technology—which requires manual configuration that must be updated each time a form is changed—by accurately extracting text, forms, tables, and other data without the need for any manual effort or custom code.

With **Amazon Textract**, you can quickly automate activities across complex documents such as loan applications, tax documents, enrollment forms, or medical claims. Once the information from these documents is captured, you can move into the next stage of your business process faster—leading to more productive employees and happier clients and customers.

KABBAGE

Saving nearly a million jobs through faster business loans

Kabbage, a data and technology company in the financial services sector, currently provides over 500,000 small businesses with access to more than \$16 billion of working capital. The Paycheck Protection Program (PPP), a component of the US government's response to economic struggles related to the COVID-19 pandemic, presented Kabbage with a unique opportunity. But it also offered a challenge: Kabbage had never issued a small business administration (SBA) loan before, and it needed a fast, automated way to process PPP applications.¹



*Amazon Textract helped support 80% of Kabbage's PPP applicants to receive a fully automated lending experience and reduced approval times from multiple days to a median speed of four hours. Kabbage became the second largest PPP lender in the nation by application volume...serving over 297,000 small businesses and preserving an estimated 945,000 jobs across America."*¹

Anthony Sabelli, Head of Data Science, Kabbage



Transform documents into usable data

Amazon Comprehend is a natural language processing (NLP) service that uses machine learning to find insights and relationships in text. The service identifies the language of the text, extracts key phrases, places, people, or brands, understands the sentiment in text, and automatically organizes a collection of text files by topic. You can train Amazon Comprehend to identify entities relevant to your organization, categorize documents, and/or assign relevant labels to text.



Our applications use Amazon Textract and Amazon Comprehend in conjunction with our own proprietary models to automate costly manual processes like document review and insurance application intake. They reduce customer effort, making it faster and easier to buy and sell commercial insurance. We are seeing great results using AWS machine learning."

Colin Toal, Chief Technical Officer, Chisel AI²

CHISEL AI

Automating manual processes for commercial insurance documents

The commercial insurance industry operates using processes that are document intensive. Processing these documents requires manual entry, is costly, and prone to human error. Chisel AI uses AWS services to automate and streamline operational workflows, using Amazon Textract and Amazon Comprehend. Using machine learning, they have been able to extract policy numbers, expiration dates, and many more insurance industry-specific attributes with reduced manual effort. They leverage Amazon Textract to extract data from documents at scale and Amazon Comprehend to classify and label unstructured content in documents and pull out insurance-specific entities.²

² <https://aws.amazon.com/comprehend/customers/>

Extract potentially lifesaving insights from medical text

Amazon Comprehend Medical extends the power of NLP to healthcare and life sciences organizations. The service makes it easy to use machine learning to extract relevant medical information from unstructured text. Using Amazon Comprehend Medical, you can quickly and accurately gather critical data—such as medical condition, medication, dosage, strength, and frequency—from a variety of sources, like doctors' notes, clinical trial reports, and patient health records.

Amazon Comprehend Medical can also link the detected information to medical ontologies, such as ICD-10-CM or RxNorm, so it can easily be used by downstream healthcare applications.

FRED HUTCHINSON CANCER RESEARCH CENTER

Using IDP to accelerate the development of cancer treatments

Fred Hutchinson Cancer Research Center is an institute with the lofty goal of curing cancer by 2025—and it's using Amazon Comprehend Medical to help make its vision a reality. The organization needed a better method of using document data to develop clinical trials and connect researchers with the right patients faster. Amazon Comprehend Medical helped provide the solution, reducing the time needed to sift through and label unstructured clinical data from hours to seconds.



“For cancer patients and the researchers dedicated to curing them, time is the limiting resource. (Deploying Amazon Comprehend Medical) is a vital step toward getting researchers rapid access to the information they need when they need it so they can find actionable insights to advance lifesaving therapies for patients.”³

Matthew Trunnell, CIO, Fred Hutchinson Cancer Research Center



³ <https://aws.amazon.com/comprehend/customers/>



Facilitate faster, more efficient human review

Due to regulatory requirements and other unique business needs, many organizations must still rely on manual processes to analyze documents in certain use cases, such as sensitive documents in healthcare, handwritten documents, or insurance claims. This generally does not mean that IDP cannot be deployed at all—only that manual review of the work done by IDP systems is required at some level.

Amazon Augmented AI (Amazon A2I) makes it easy to build and manage human reviews for IDP and other machine learning applications. The service provides built-in human review workflows for common use cases, such as content moderation and text extraction from documents. Using Amazon A2I, you can send any document to a human for review to ensure the text, phrase, or information is processed correctly. Additionally, you can use this human review information to retrain your machine learning model to provide for accuracy downstream.



NATIONAL HEALTH SERVICE

Integrating machine learning and human judgment in healthcare

National Health Service Business Services Authority (NHSBSA) provides a range of support services to healthcare organizations, contractors, and patients in the United Kingdom. As part of those services, NHS processes 54 million paper prescriptions per month. The organization requires an IDP solution that can do that work quickly—but can also enable fast, easy human intervention where necessary.⁴



We are excited about Amazon Augmented AI because it allows us to take advantage of machine learning while still applying human judgement. That's a game changer for us."

Chris Suter, Head of Cloud Platforms and Innovation, NHSBSA

⁴ <https://aws.amazon.com/augmented-ai/customers/>





Prior to our deployment of Kendra, we attempted to create logical navigation paths to guide our customers to the information they needed; however, none produced the desired level of customer experience (CX) and could lead to frustration or worse, customers making less than optimal decisions about their health benefits. Amazon Kendra plugged in effortlessly to our existing portal platform and freed up staff members to focus on higher value activities to drive CX improvements. While we're still early in our use of Amazon Kendra, we've already seen it support thousands of positive customer interactions and deliver higher quality answers in just the first few days."

Devin Parsons, VP and Head of Digital Transformation,
bswift, a CVS Company

Find accurate and relevant information, faster

Nearly half the time, users fail to find the information they need to excel because information is scattered across their organization in the form of documents and other unstructured data. This negatively impacts customer experiences as well as workforce productivity.

Amazon Kendra delivers intelligent search and natural language understanding capabilities, enabling employees and customers to ask natural language questions to find relevant content contained within documents. Amazon Kendra will quickly surface relevant content as a suggested answer instead of a list of links, so employees get the information they need when they need it.

BSWIFT

Making enterprise knowledge searchable through machine learning

Since inception, bswift has focused on using technology to simplify the administration of healthcare, reduce costs, and connect customers to the resources they need to achieve their health ambitions.⁵

⁵ <https://aws.amazon.com/kendra/customers/>

Amazon SageMaker

Organizations that want to develop their own machine learning models for text extraction and analysis can use [Amazon SageMaker](#), a fully managed service that helps data scientists and ML developers build, train, and deploy machine learning models quickly. For text extraction, Amazon SageMaker offers built-in algorithms, such as BlazingText, pre-trained models available through the AWS Marketplace, and the ability to develop your own text-processing algorithms. Regardless of which option you choose, Amazon SageMaker provides all the tools you need for machine learning end-to-end—so you can easily develop high-quality text-processing models.

SIEMENS FINANCIAL SERVICES

Building custom IDP solutions with Amazon SageMaker

Siemens Financial Services provides business-to-business financial solutions, offering equipment, transport, and automobile leasing services. It uses an NLP model it developed with Amazon SageMaker to extract information from documents and accelerate investment due diligence—reducing time to summarize documents from 12 hours to 30 seconds.⁶

THOMSON REUTERS

Helping customers find the information they need

Thomson Reuters, the world's leading source of news and information for professional markets, designed an NLP capability in the context of a question-answering application, which delivers accurate information to customers and allows them to simplify and derive more value from their work.⁷



⁶ <https://pages.awscloud.com/NAMER-In-GC-400-machine-learning-sagemaker-tco-learn>

⁷ <https://aws.amazon.com/sagemaker/customers/>





AWS IDP solutions for specific industries and use cases

AWS IDP solutions use machine learning to help organizations of all sizes and industries extract and analyze document data faster. But in this section, we'll identify and explore a few industries and use cases where the technology is particularly beneficial.

Financial services organizations

Documents, such as mortgage applications, at financial services organizations contain sensitive information that must be analyzed with absolute precision. The smallest of errors could result in the wrongful denial of a loan or a costly overpayment of interest.

With Amazon Textract, financial organizations can extract text, forms, and tables from documents. They can then use Amazon Comprehend to analyze entities, phrases, or key-value pairs within the extracted text and Amazon Augmented AI (Amazon A2I) to facilitate manual review of sensitive or nuanced results. For example, an organization deploying all of these services could pull key phrases—like company names, social security numbers, and interest rates—from documents using Amazon Textract, analyze the information to discover patterns and opportunities with Amazon Comprehend, and perform manual accuracy checks through Amazon A2I. They can then make this extracted and analyzed content searchable throughout the organization using Amazon Kendra.

Healthcare organizations

By deploying both Amazon Textract and Amazon Comprehend Medical, healthcare organizations can quickly and easily extract data from tables within documents—and then group key-value pairs or entities to understand relationships within the document. Healthcare organizations use these solutions to gain a more holistic view of data spread across various documents. For example, a provider or researcher can analyze symptoms from multiple documents and/or patient histories to better understand how particular symptoms relate to a specific diagnosis.

Contact centers and customer support teams

Organizations across nearly every industry can use AWS IDP services to augment capabilities for their contact centers and customer support teams. They can use Amazon Comprehend to extract and analyze important text from emails and support tickets. Amazon Comprehend can also help you find positive or negative reviews and identify features that are commonly mentioned when customers are happy or unhappy with the business. Having faster and more comprehensive access to this information will help increase understanding and incorporate feedback to improve customer experiences, products, and services.



Free critical data

Ready to begin your IDP journey? AWS provides a range of options that can help you **start realizing the benefits of machine learning-powered IDP today.**

You can leverage the **AWS Professional Services organization**, a global team of experts that can help you deploy IDP and other services to realize your desired business outcomes when using the AWS Cloud.

Or, you can **train your developers** and data scientists to build custom IDP models and gain a stronger understanding of machine learning in general. Our training initiatives use the same curriculum we use at Amazon, and many courses are available on demand and at no cost. We can help everyone in your organization—executives, developers, and data scientists alike—become more proficient in machine learning.

Finally, you can **contact us** directly for more information on IDP and other machine learning solutions for your organization.

No matter what path you choose, one thing is certain: to remain competitive in today's business world, you can't afford to continue to use manual, time-consuming methods of document processing. By setting your data free and transforming it into actionable insights with IDP solutions from AWS, you can boost productivity, cut down on costs, and improve customer satisfaction.

Explore more ways IDP can benefit your business ›

