IRON
MOUNTAIN®

POWERING YOUR CONTENT WITH AI

A 10-MINUTE GUIDE TO ARTIFICIAL INTELLIGENCE IN CONTENT SERVICES PLATFORMS

CONTENTS

INTRODUCTION

With artificial intelligence (AI) now becoming one of the hottest
topics in technology, we are at an unprecedented time in the area of
information management and, in particular, Content Services. Never has
a particular technology held so much promise and, yet so much hype
and misunderstanding.

This "10-minute" guide, explores the rapidly evolving role that AI and
machine learning (ML) technologies play in content management. The
paper reviews some available AI offerings and their practical application
for enabling better access to critical information. Discover real-world
use cases and how early-adopters get business value out of these
technologies. This guide will help you understand the critical enterprise
considerations for getting started with AI and Content Services.

THE CHALLENGE OF CONTENT AND THE PROMISE OF AI

Content – in all of its various forms – has long been a challenge from an information management perspective. It can be nearly impossible to find due to inadequate and inconsistent metadata attributes, limited search functionality within core business applications, and disconnected repositories/systems.

In a recent survey of UK financial services companies, 80% of respondents indicated that their systems were not fully integrated and their organizations had an average of nine different content management systems in place. Further, employees at these organizations spend almost one hour, every day, simply looking for the information that they need to do their jobs – that's a 15% loss in productivity.
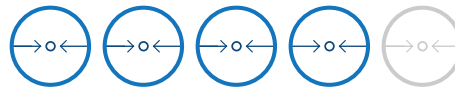
Organization's systems are not fully connected to each other

Time spent daily looking for information   52 min.

Average number of solutions

60%   **See the potential of AI to automate mundane tasks**

60%   **Believe their organization lacks the skills to capitalize on AI**

Not only is content inherently hard to manage, the volume and types of content are growing at an unprecedented rate. Many enterprise organizations have accumulated over 10 billion documents and scanned images over the last 20 years – an impressive amount of content. But today, they're looking to capture and manage in excess of 500 million new objects per month. They will literally double their entire corpus of content in the next two years.

Content, in general, isn't difficult for a human to understand. We consume content every day without even thinking about it. The challenge is that the things we do naturally – quickly classifying content to determine what it is, identifying critical information and data within it, and perhaps determining if it's vital corporate information that needs to be preserved – don't scale well.

Extracting information from content and entering it into fields and tables is work that people inherently don't like to do. And doing this work across 1000s or even 100s of 1000s of new documents, every day, is challenging, expensive, and difficult to do with consistent accuracy. This is why so many organizations have struggled with enterprise content management (ECM) for so long.

Now, with the advent of AI and machine learning, we've found a way to process content like a human does, but at a massive scale. We can use a range of services to extract critical data from content and, in doing so, transform content into intelligent information that can be easily found, readily used to perform work, and accessible any time, anywhere, and on any device.

PUBLIC CLOUD AI SERVICES

Most modern Content Services Platforms can integrate with a variety of public cloud services for artificial intelligence. Typically, the Content Services Platform will pass an object – a document, image, or even a video file – to a cloud provider and will receive a set of data produced by the AI service.
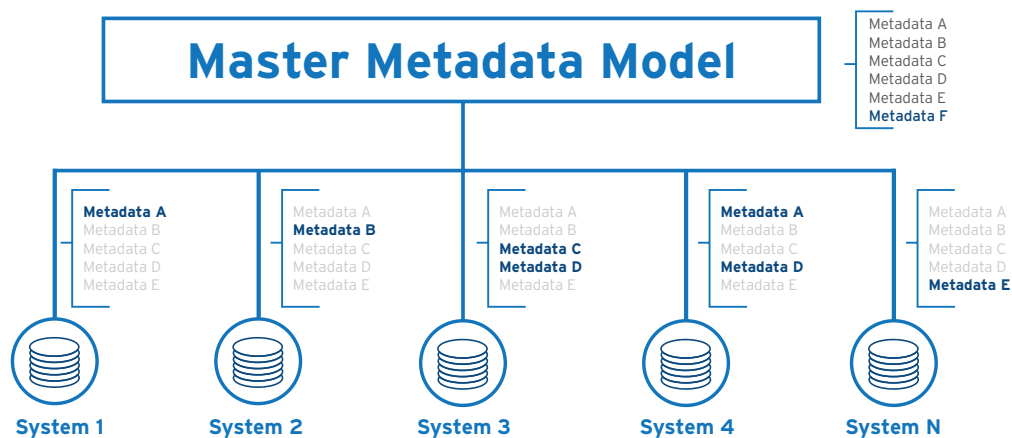
The world of AI continues to evolve rapidly, and a number of large technology companies now offer a variety of commodity AI services that can be leveraged for working with various forms of content. Let's quickly explore some of the most popular public offerings and some examples of how they can be employed to work with content:

The world of AI continues to evolve rapidly, and a number of large technology companies now offer a variety of commodity AI services that can be leveraged for working with various forms of content. Let's quickly explore some of the most popular public offerings and some examples of how they can be employed to work with content:

AMAZON COMPREHEND

A Natural Language Processing (NLP) service that employs machine learning to perform entity extraction, sentiment analysis, and language detection on text. It can also perform document classification.

Example use case: perform sentiment analysis on customer emails and chat session content to immediately identify unhappy customers for priority response and faster resolution
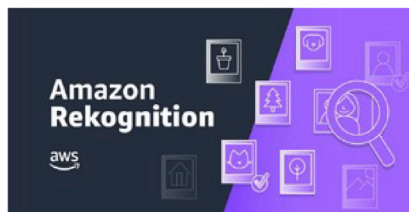
AMAZON REKOGNITION

A deep-learning technology to identify objects, text, people, scenes, and activities in videos and images. Rekognition can also detect inappropriate content and perform facial recognition.

Example use case: identify celebrity images used in advertising content
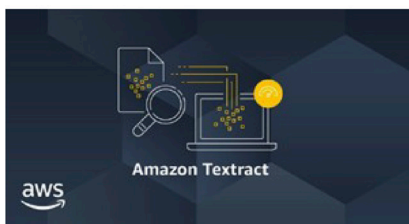
AMAZON TEXTRACT

A machine-learning service to perform Optical Character Recognition (OCR) on scanned documents (images) and to extract specific data values.

Example use case: forms recognition and processing

AMAZON TRANSLATE

A neural-machine service for language translation.

Example use case: automatically translate sales and marketing collateral into various local languages

AMAZON TRANSCRIBE

A deep-learning process to convert speech to text quickly and accurately.

Example use case: transcribe customer service calls to text which can then be processed with sentiment analysis to identify unhappy customers

GOOGLE VISION

A machine-learning service that classifies and assigns labels to images, detects embedded objects and faces, and even extracts text.

Example use case: read license plates in an automobile accident photo

GOOGLE DOCUMENT AI

A machine-learning service to perform text, character and image recognition in 200 different languages. Document AI also provides sentiment analysis, entity extraction, and other NLP capabilities.

Example use case: sentiment analysis on emails or social media content

MICROSOFT COGNITIVE SERVICES

Microsoft offers an extensive set of machine-learning models under its Cognitive Services. These include speech to text, text to speech, computer vision, forms recognition and text analytics.

Example use cases: OCR, forms processing, sentiment analysis, etc.

Many of these machine-learning offerings are focused on providing greater insight into and understanding of content, whether that's text- based documents, photos and images, or even audio and video files.

A lot of value can be derived from these generic models and services, particularly in performing routine tasks with high volumes of content. For example, if you need OCR for a large existing set of content, these services are accurate and highly performant. Real-time sentiment analysis of chat sessions, emails or even social media content is another great use case for these services.
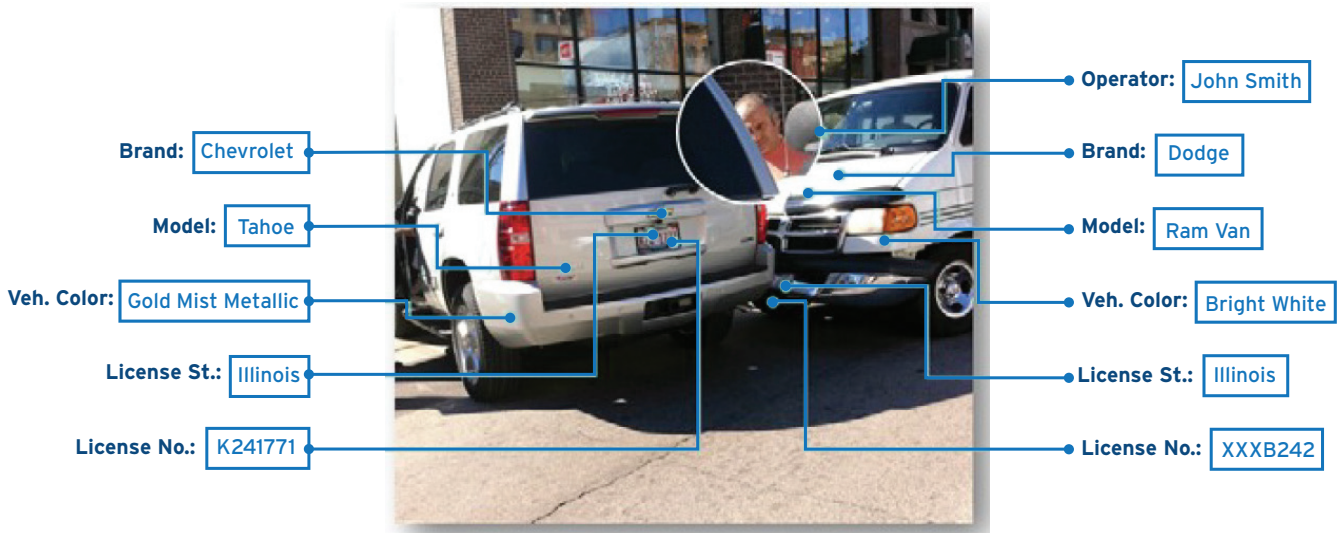
GENERIC VS CUSTOM MACHINE-LEARNING MODELS

Generic services have been trained with a broad range of data and, as a result, these models tend to return generic data, which may or may not be helpful depending on the use case.

Here's an example. The picture below shows an automobile accident which has been labeled by Google Vision.
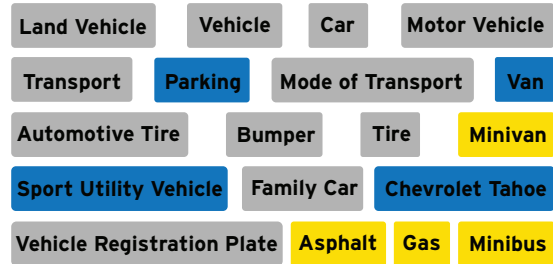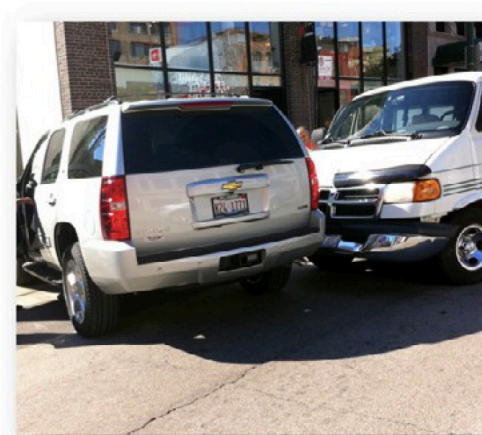
As you can see, Google Vision has returned a number of labels or data values related to the image, however if you were an automobile insurer, is this data really valuable?

One of the real benefits of machine learning is that organizations don't have to rely on generic models and generic data. They can use machine learning to train their own, custom models that will return data that's much more specific to the needs of their business.

Now, let's consider a machine-learning model that has been specifically trained with lots of pictures and data related to automobile accidents. Above is an example of the kind of data that a custom model could extract from the same image.

Note that now the make, model, and factory color of both vehicles have been correctly identified. We have also identified two Illinois license plates and captured full and partial plate numbers. There is a face present in the image and identified as the operator, Jim Smith.

Assuming this picture was taken with a smartphone or other digital device, it's likely we could also use the GPS coordinates from the picture and identify the accident location. No, this isn't AI, but probably useful for processing the claim.

Instead of returning a generic set of labels for an image, a custom model enables you to perform entity extraction.

The benefits of a custom model include:

1. Extracting "business specific" data that adds real value

2. Bringing greater automation to the process

Not only are we no longer dependent on a human to enter these values, we can also automatically alert a claims processor that new information is available for this claim. Content and data coming together to ensure the right information is in the right place at the right time.

LET'S GET REAL

ENRICHMENT

Enrichment is all about extracting data from content and using this data to make that content more accessible, more contextual. Enrichment can take on many different forms, depending on what type of content you're working with.

In a Content Services Platform, data provided by custom machine learning models gets applied to the object – an image, video, document, or other content type – as metadata which is indexed and can later be used to find and retrieve that object. That data can also be used to trigger workflows and initiate processes. It can also be passed to other, integrated systems. In the insurance example, this may be valuable claims data that needs to be recorded in a claims processing system, like Guidewire or Duck Creek.

Machine-learning models can also be used to enrich traditional content, like documents and scanned images.

For example, many financial services firms have large volumes of existing TIFF images that they want to convert into PDF documents. The benefit of converting TIFF images into PDF Documents is that PDF documents can be indexed and searched based on the entire context of the document, commonly referred to as full-text indexing. Users can search for content within a PDF document, perhaps to identify a particular word, phrase, or even a contractual term.

In order to accomplish this, OCR must be performed on the TIFF image to extract all the text which is later indexed. (A public AI service, like Amazon Textract, can be used to OCR the content.) A transformation service (not AI) is used to map this text back to the original image and convert the image to a PDF document. This document is then ingested into the Content Services Platform and properly indexed for search and access.

AUTOMATION

AI and machine learning can help companies better automate critical business functions and processes.

Example One: Data Validation

> A number of enterprise organizations still process millions of paper forms every year, the majority of which are handwritten. A critical challenge is to first determine if the form has been completed correctly, before processing it. Until now, this has been a labor- intensive and therefore expensive process. A human is required to determine what type of form it is, and then validate that the necessary responses have been provided, signatures are in the right place, and if required, confidential customer information is present.

Machine-learning models can do all of this formerly manual work. Organizations can capture these forms from a variety of sources – fax, email, or physical forms – and convert them into digital images/documents. Machine learning is then used to identify the different forms and perform the necessary validation on the provided information, in most cases before a knowledgeable human even looks at the document.

Machine-learning models also enable intelligent exception management, to quickly identify what's missing or wrong with a provided form and automatically route it to a customer service representative or back to the customer for remediation.

Example Two: Records and Retention Management

> For years, many organizations have struggled to implement an effective records management approach for their information. The simple reason is that most organizations are unwilling to devote the effort required to look at all of their existing content to determine if, when and how it should be retained or even deleted.
>
> This is painstaking work for a human, but machine learning can automatically classify content and extract data from it at extreme scale.
>
> As a result, it's much faster and easier to examine tremendous amounts of content, classify the different types of documents or information, and then automatically identify records, apply requisite retention periods, and even delete non-vital information.

INSIGHT

Deriving new insights and intelligence about business content enables digital transformation.

Example One: Fraud Detection

> Unfortunately, in insurance, fraudulent claims remain prevalent and it is not uncommon for claimants to use the same accident photos in multiple claims.
>
> If you think about the average large P&C insurer, with thousands of claims processors, what are the odds that the same claims processor is going to handle two claims with the same accident photos? And perhaps even photos that are resubmitted years apart? Not very good.
>
> AI can help detect insurance fraud by leveraging machine learning to compare new claims photos with existing photos, quickly identify duplicates, and then automatically launch a fraud investigation process.

Example Two: Repair Estimates

> Insurers can use machine-learning models to identify claims with similar accident damage in vehicles that are of the identical make, model, and year.
>
> The insurer can then cross-reference actual repair costs to come up with a real-time estimate or range for the damage depicted in the photograph. This information could be immediately shared with the insured (customer) and also used to validate estimates that were subsequently submitted by the insurer's repair network, thus automating repair approvals.

Both of these are great examples of how organizations can leverage existing sets of data and content – what we might term content lakes- to add new value and new levels of intelligence and insight to their businesses.

ENTERPRISE CONSIDERATIONS

Now that we have explored the difference between commodity and custom models and have examined some real-world use cases for AI/ ML and content, let's look at some key considerations for organizations that are considering a Content Services Platform with enterprise-class AI/ML capabilities.

COMMON INTEGRATION FRAMEWORK

A Content Services Platform should not only provide its own discrete services for content, it should also readily plug into external services, like commodity AI services from providers like Amazon, Google and Microsoft.

A best practice is to leverage a common integration framework for external AI/ML services. This approach is highly adaptable as the framework can be easily configured for different content structures and work processes. This will also provide a common method to access these services across the platform. It should be extensible to provide a standard framework for future integrations as new AI service offerings become available. And, of course, it should be highly scalable and performant – to support large volumes of content and data – as well as resilient and recoverable.

As you consider your options for AI and Content Services, look for integration approaches that were designed for enterprise use cases.

SUPPORT FOR CUSTOM ML MODELS

Many organizations have begun experimenting with AI and may even have highly skilled Data Scientists on staff. Others are just beginning their journey. If you are an organization that is considering deploying your own custom, machine-learning models, it's very important to know where you are in this journey and to choose a Content Services Platform (CSP) that's flexible in its approach to AI/ML.

For inexperienced organizations, your CSP should allow you to easily:

> Configure a new custom model to identify what particular data you want to extract

> Identify and export training sets from your existing content repository, perhaps even providing recommended training sets

> Import and deploy newly trained models into production

> Use wizard-driven interfaces and point-and-click configuration

For experienced AI practitioners, your CSP should be:

> Easily integrated with in-house systems to publish ML models and expose existing content and data training sets

> Highly scalable export systems that can be easily configured to export ML training sets

> Built around open standards – like Tensorflow – that allow you to deploy your own, in-house-developed ML models and even supply your own custom training and pre-processing algorithms

CONTINUOUS TRAINING & ADMINISTRATION

Another critical consideration is how your ML models perform over time.

First, you should consider solutions that utilize continuous training paradigms that enable your ML models to evolve and improve over time as new content and data is added to the system. Human interaction with machine-generated data is also critical to provide data validation and further train ML models. Look for a Content Services Platform that considers the human role in the machine-learning process and provides specific interfaces for "human in the loop" training.

Your Content Services Platform should also provide real-time performance monitoring for models. ML models can begin to show bias or even degraded performance, therefore, a performance monitoring interface will help identify models that have become corrupted or are showing degradation in performance. Machine-learning models should also be versioned, allowing you to quickly roll back to an earlier version, should your model become degraded.

AI GOVERNANCE

Artificial intelligence is a new frontier and, for now, regulation and compliance are trailing far behind the technology. It's important for organizations to begin thinking about proper governance for AI/ML.

Look for a Content Services Platform that can differentiate machine-generated data from human-generated data and corrections. Consider solutions that can rollback ML models to previous versions and also rollback all of the data that model may have generated.

We also recommend a Content Services Platform that stores the training and evaluation datasets that were used to create different models. In this circumstance, an organization can not only identify which model created which data, they can also represent how the model was originally trained to produce that data.

GETTING STARTED

We hope that this has been a valuable introduction into the world of AI, Machine Learning and Content Services Platforms. At this point, you might be saying, "This is all great, but how do I get started?" We'd like to help.

Visit to learn more about Iron Mountain InSight's modern, cloud-native Content Services Platform and be sure to read about our InSight service and how we're making it easy for InSight customers to train and manage their own custom ML models.

IRON MOUNTAIN®

1300 476 668 | IRONMTN.COM.AU

0800 732 255 | IRONMOUNTAIN.CO.NZ