# Employing AI for Retention & Disposition in Digital Information and Recordkeeping Systems

## Inventory of AI-enabled Software and Services

InterPARES TRUST AI

The following members of this study contributed to the completion of this *Inventory of AI-enabled Software and Services* report.

**Lead Researcher:** Dr. Patricia C. Franks

**Project Researchers:**

| | |
|---|---|
| Georgia Barlaoura | Isto Huvila |
| Jason R Baron | Pat Moore |
| Jenny Bunn | Olefhile Mosweu |
| Lois Evans | Sarah Weldon |
| Prisca Giordani | |

**Graduate Research Assistants:**

| | |
|---|---|
| Melissa Kemp (SJSU) | Jordan Kerr (UBC) |

# Employing AI for Retention & Disposition in Digital Information and Recordkeeping Systems

## Inventory of AI-enabled Software and Services

## 1. Introduction

The biggest challenge facing archivists and records and information professionals today is how to deal with big data which possesses the characteristics of volume, velocity, variety, veracity, value, and variability. Big data is so massive it can't be controlled using today's database tools and analytics software techniques. It requires advanced technologies and techniques to automate routine processes. Even simple automation tools and workflows are not enough. Vendors are increasingly marketing their AI-enabled software and services to aid information professionals.

## 2. Purpose and Scope

Regardless of the records lifecycle model an organization follows, all records go through the following areas: records capture and classification; records retention and disposition; records integrity and maintenance; and records discovery, use and sharing. Decisions made in any phase of the lifecycle will impact the ability to perform retention and disposition tasks. In addition to the term "records," organizations may use other terms (e.g., data, content, information) to identify business assets of value. To understand how existing software and services use Artificial Intelligence technologies to capture, manage, and preserve digital content, a subgroup of the *Employing AI for Retention & Disposition in Digital Information and Recordkeeping Systems* research group embarked upon a preliminary inventory of commercial applications. The overall goal is to understand how AI tools and techniques have been or are being introduced into commercial Digital Information and Recordkeeping Systems

## 3. Methodology

The following steps were taken to review the software applications included in this report.
   a. All members of the subgroup recommended vendors/products for inclusion by adding them to a worksheet.
   b. Members then selected those vendors/products they wished to explore further.
   c. Data-gathering documents were prepared by researchers with the goal of identifying AI-enabled features.
   d. Based on data discovered, the software applications were ranked according to their application to this study and their use of artificial intelligence/machine learning technologies: 3 – yes, 2 – not at this time, 1 – no.

    e.   All software applications are included in this report but categorized into 3 sections.

    f.   Graduate research assistants and the study lead reviewed, revised, and reformatted the data-gathering document for the report.

## 4. Terminology

The basic reference for terminology for all InterPARES AI projects is the InterPARES Trust AI Terminology Database found at [https://interparestrustai.org/terminology](https://interparestrustai.org/terminology). As described on the website, "the *InterPARES Trust Terminology* is a vocabulary drawn from the emerging and evolving intersection of recordkeeping and information technology. It continues the work of previous InterPARES projects." If a term, such as Artificial Intelligence, is not yet found in the InterPARES Trust Terminology database, researchers refer to a working copy of terms to be added. If terms are not available within the terms to be added, ISO documents are then referred to. The main ISO documents referred to for this report are:

- ISO 15489-1:2016
- ISO 23801-
- ISO 16175
- ISO 30300

## 5. Literature Review

A literature review was prepared for this study, *Employing AI for Retention & Disposition in Digital Information and Recordkeeping Systems (AA01).* The literature review completed in May 2022 provided a brief overview of artificial intelligence, machine learning, and algorithms and how they work. The literature review then reviewed articles related to the Records and Information Management (RIM) Lifecycle, which demonstrated that appraisal and classification tools could be used or adapted to determine how long items or collections should be retained and assist in the disposition process if no longer of value to the organization. A model need not have been created specifically to complete retention and disposition tasks to be utilized for the purpose. From custom-built tools to commercial e-discovery and software-as-a-service tools, various artificial intelligence tools are being used or could be explored to aid in retention and disposition in Digital Information and Recordkeeping Systems. Those findings prompted the formation of the subgroup to explore commercial products and services that employ AI for some portion of the RIM lifecycle.

## 6.  Findings

AI has a range of uses in business and government, especially for records management and archival needs. It can help users accomplish more within tight budgets, automate tedious tasks, enhance mission-critical capabilities, design better policies, and make better decisions. The objective of this report is to provide examples of AI-enabled

products and services on the market today and to share information gathered to provide readers with insight into what our team discovered that may help them when considering the implementation of AI tools and techniques into their operations. Through an iterative process, the initial recommendation of 56 vendors/products was reduced to 41. Results of a further review of the 41 vendor products/services led to 25 that have been recommended for the interview stage of the study.
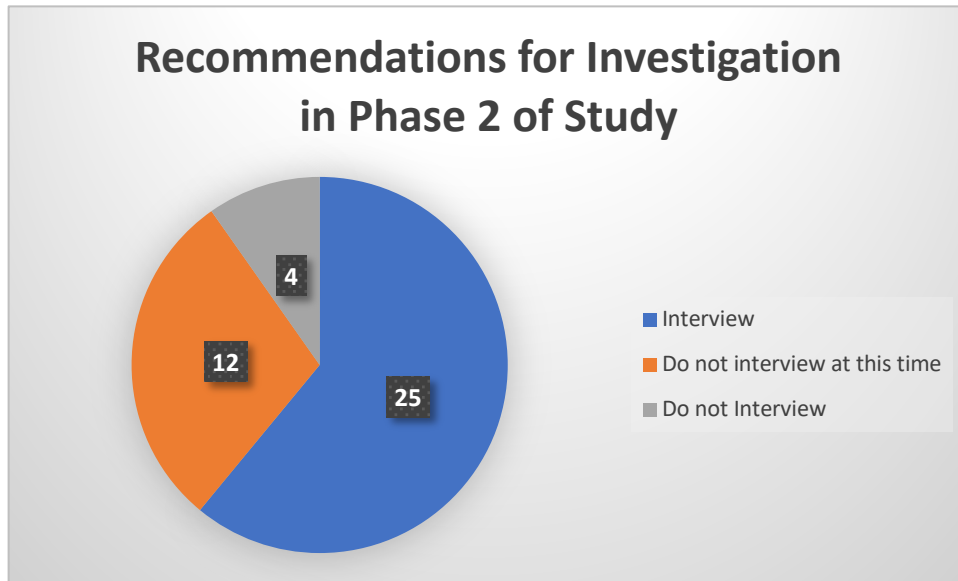


**Figure 1.** Breakdown of recommendations based on an initial review of 41 vendor products/services.

The main contents of this report are divided into 4 sections: 1) 25 vendor products/services recommended for interviews, 2) 12 vendor products/services of interest but not recommended for interviews at this time, 3) 4 vendor products/services out of scope of our study, and 4) a summary of the findings.

## 7. Collaboration with CU05 Study

The members of the AA01/SG2 team, led by Jenny Bunn, have been meeting with the members of the CU05 study team led by co-chairs Mariella Guercio and Stefano Allegrezza. Their study, *The role of AI in identifying or reconstituting archival aggregations of digital records and enriching metadata schemas*, is also exploring the status of the use of AI in commercial products and services. To avoid duplication of efforts, the AA01 team will interview only those vendors CU05 has not already committed to interview. Of the 25 AA01 has identified, CU05 will interview 11. AA05 will interview the remaining 14.  To ensure standardization of data obtained through the interviews, the AA05 team will use the Questionnaire designed by the CU05 team. Joint products will be determined, such as a joint publication of the findings will be released, and potential software demonstrations of the most useful products will be discussed. The tentative schedule, developed in Nov. 2022, is shown in Table 1.

**Table 1.** Timetable of tasks for cooperation developed in November 2022.

| Date | CU05 | AA01 |
|---|---|---|
| End of Dec 2022<br>We agreed to check back with our respective teams to see if this idea met with approval and whether the proposed timetable would work and fit in with our respective schedules. | Complete Current data collection | Prepare to approach additional vendors in Jan 2023 |
| Jan-Apr 2023 | Initial data analysis | Data collection following same approach |
| May-July 2023 | Informed by CU05's findings – collective data analysis | |

Following this or a similar timetable the goal is to bring all the data together in the early summer of 2023 to create the final output and/or possibly lead into following up with companies to arrange demonstrations.

## 8. Next Steps

The members of the subgroup charged with exploring the use of AI tools and techniques in commercial products and services will continue their work as reflected in Table 1 in Section 7. The findings of this work will be incorporated with the work of several other subgroups within the larger AA01 study.

# Inventory of AI-enabled Software/Services
## Ranking #1
## Recommended for Interviews

# AI Enabled Services and Software

| | |
|---|---|
| **Name of Vendor:** | ActiveNav |
| **Name of Product/Service:** | ActiveNav Software and Services |
| **Self-Described Service:** | Sensitive data identification and content cleanup at scale. Providing capability for defensible deletion of ROT. |
| **Website Address:** | https://activenav.com/ |
| **Gartner CSP Magic Quadrant:** | Yes / No |
| **G2 Grid for ECM:** | Yes / No |

**Introduction:** ActiveNav is a small company based in the U.K. and with a large footprint in the US, with clients consisting of various US and Canadian civil federal agencies, components of the Department of Defense, and Fortune 1000 companies. The core of ActiveNav's business is providing actionable insight into data repositories for the identification and mitigation of regulated sensitive data risk as part of data clean up and ongoing governance program. Additional use cases include identification of intellectual property, record disposition, and the removal of redundant, obsolete, or trivial (ROT) data.

| AI Features |
|---|
| **AI techniques/approaches:** <br> ActiveNav's released products currently primarily use rule-based approaches to support classification of data and proprietary object scoring; for some use cases natural language processing is used to identify key phrases and to summarize document content. <br><br> ActiveNav is running R&D projects to assess the effectiveness of automated classification based on machine learning techniques. <br><br> ActiveNav's Data Mapping-as-a-Service compliance tools utilize AI to provide a holistic and real-time view of the organization's data universe. <br><br> ActiveNav's cloud-hosted data discovery service combines a range of AI capabilities and visualization tools to deliver these capabilities at scale across a wide range of industry verticals. |
| **What functionality does AI enable:** <br> Rule based classification approaches allow labeling of content of interest for a range of use cases. For instance, identifying redundant, sensitive, or risky data. These findings are then visualized within the product to support migration and cleanup activities. <br><br> Additional use cases: data classification and scoring; targeted search for Data Subject Access Requests (DSARs) and Freedom of Information Act requests; extraction of features and data elements from enterprise-scale volumes of unstructured data (text, tabular, and image); identification and visualization of data hotspots (e.g., for sensitive data risk analysis). <br><br> ActiveNav envisions introducing machine learning based approaches to support users in the review of initial automated findings, utilizing their subject matter expertise to build classifiers that can accelerate the task of assessing findings.[1] |

---

[1] https://siliconangle.com/2021/09/21/active-nav-debuts-data-mapping-service-compliance-tools/

# AI Enabled Services and Software

**Software Offerings (AI-enabled features):**
- Rule-based classification is based on underlying pattern-based feature extraction supported with Boolean search rules.
- Natural language processing uses a proprietary methodology using a hybrid linguistic/statistical technique.
- Model-based classification based on viewer behaviors.
- Semantic ontology OEM services.

**How are functions enabled?**
- Rule-based classification is based on underlying pattern-based feature extraction supported with Boolean search rules.
- Natural language processing uses a proprietary methodology using a hybrid linguistic/statistical technique.
- Object scoring uses a proprietary statistical model.
- Predictive review is based upon models built from the analysis of users' prior review behavior.

**What descriptions of algorithms are provided?**
None

**Third party services supporting AI?**
None specified.

**Opportunity for User input? If rule-based, can it be customized?**
- Customers are provided with standardized models for rule-based classification. These can then be refined based on specific requirements and the characteristics of their data.
- They plan for initial capabilities using machine learning techniques to build customer specific models based on user activity.

**What information about the AI models in use is provided/available to the user?**
- Rule based classification models are open for review and refinement.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
As explained above.

**Assessment:** An essential approach to managing content/information is to manage only that which has value to the organization.  One challenge is to discover where all content resides and then create a data map to monitor that data across all applications like Microsoft SharePoint, email, chat and social media through the use of a data map. ActiveNav applies AI to provide Data-mapping-as-a-service to provide the organization with an accurate, holistic view of their data risk profile.  Identifying and defensibly deleting e-trash, ROT (redundant, obsolete, and trivial), stale, and low-value content optimizes data storage infrastructure is a logical next step.

**Recommendation:**  The AA01 study group *recommends* learning more about ActiveNav's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

**Name of Vendor:**              Adlib
**Name of Product or Service:**   Document Transformation Platform
**Self-described Service:**       Document conversion and transformation platform
**Website Address:**              https://www.adlibsoftware.com/platform
**Gartner CSP Magic Quadrant:** Yes / No
**G2 Grid for ECM:**             Yes / No

**Introduction:** Adlib's Document Transformation Platform is an AI-drive cloud solution that leverages artificial intelligence (AI), machine learning (ML), and natural language understanding (NLU). It transforms unstructured document data locked in documents across emails, file shares, and repositories into intelligent data to be used by technology to maximize the potential of people and processes.[1]
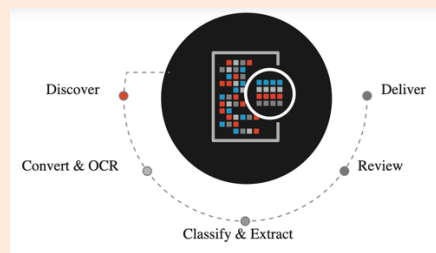
| AI Features |
| --- |
| **AI techniques/approaches:** Natural language processing and machine learning |
| **What functionality does AI enable:** <ul><li>Detect duplicates</li><li>Conceptualize (discover, assess, prepare): discover file metadata and apply "file analytics" to identify redundant, obsolete, and trivial data for remediation;</li><li>Standardize (normalize, recognize, optimize): capture documents, covert over 400 file types to PDFs with "advanced" OCR;</li><li>Analyze (interpret, fingerprint, cluster): organize textual information into a unique file fingerprint and automatically cluster documents to expedite classification;</li><li>Categorize (classify, detect, extract): classify documents to taxonomy [i.e., manually assign a document type] and extract additional metadata [using algorithms] to strengthen the quality of the file as asset;</li><li>Visualize (view, explore, refine): provide an inventory view across unstructured data sets to aid insights and comprehension of all content.</li></ul> |
| **Software Offerings (AI-enabled features):** Document transformation, digital mailroom, content migration, supply chain management, robotic process automation, and data privacy and compliance.<br><br>AI features are employed throughout the process shown in figure 1.<br><br><br>**Figure 1:** Adlib's Document Transformation Platform. Source: https://www.adlibsoftware.com/platform |

---
[1] https://theenterpriseworld.com/the-age-of-ai-adlib/

# AI Enabled Services and Software

In the *Discover* phase, the system crawls the location and creates metadata for each file: file location; file creation date; file last referenced date; file size; file format; and file document name. This process allows for the removal of duplicate and empty files as the first stage in reducing ROT.

In the *Convert & OCR* phase, Adlib utilizes advanced Optical Character Recognition (OCR) to convert images into fully searchable PDFs, automatically publishes to PDF or PDF/A format, and renders documents as thumbnail images for users to preview contents before downloading. Intelligent and automated document assembly capabilities including merge, watermarks, and security settings are present.

In the *Classify & Extract* phase, files are run through workflows: clustering, to group like documents based on content similarity; fingerprinting, to compare, contrast, and extract information about file metadata to provide a unique, [matchable] identifier; manually classify, to assign a document type to each type of file.

**How are functions enabled?**
intelligent document processing, digital mailroom, content migration, supply chain management, robotic process automation, and data privacy and compliance.

**What descriptions of algorithms are provided?**
- Unsupervised machine learning supports document clustering where natural language processing (i.e., text extraction based on word pairs) creates document "fingerprints" that are compared, clustered, and then manually categorized by document type.

- Semi-supervised rule building leverages a labeled training set to create rules that are then applied to unlabeled content.

**Third party services supporting AI?**
None specified.

**Opportunity for User input? If rule-based, can it be customized?**
The AI models are built using the Workbench module, which allows expert users (including RMs) to create or tweak the system to automatically recognize and classify large volumes of documents.

The following information was derived from the UK National Archives' project, which investigated 4 tools out of 24 initially identified. Adlib's solution then marketed as the *Adlib Elevate Platform* was one of the four. [2] Six steps were identified by Adlib: 1) create a project; 2) create training and QA data sets; 3) build training and QA models; 4) test models; 5) review results; 6) deploy model. The Adlib Elevate models considered various attributes such as Vocabulary, Concepts, and NLP to make a classification decision. During the UK National Archives' project, some experimentation was done to assess what would lead to the best model. One of the tests that produced good results was to subsample only Word documents and create a model to run against other Word documents.

**What information about the AI models in use is provided/available to the user?**
Overall, there appears to be a level of transparency in that expert users can set up and test their own models.

---

[2] https://cdn.nationalarchives.gov.uk/documents/phase-1-market-research-ai.pdf

# AI Enabled Services and Software

> **Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
> For classification, semi-supervised rule building leverages a labeled training set to create rules that are then applied to unlabeled content.

**Assessment:** Although this product was marketed as Adlib Elevate when investigated by The National Archives of the UK (TNA), the favorable comments included in their report make further research into Adlib's Transformation Processing Platform advisable. In addition, machine learning as detailed in the description as well as a heightened ability for user input and design makes this software an attractive option for further analysis and demonstration.

**Recommendation:**

The AA01 study group *recommends* learning more about Adlib's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

**Name of Vendor:** Amazon AWS
**Name of Product / Service:** Amazon AI Services
**Self-Described Service:** Pre-trained AI Services to provide ready-made intelligence to applications and workflows
**Website Address:** https://aws.amazon.com/machine-learning/ai-services/
**Gartner CSP Magic Quadrant:** Yes / No
**G2 Grid for ECM:** Yes / No

---

**Introduction:** A range of services are offered under the umbrella label of AWS AI services. These services include entity recognition, sentiment analysis, PII Identification and Redaction, Keyphrase extraction, events detection, language detection, syntax analysis, topic modelling and classification.

Among the AWS AI services are Amazon Transcribe, Amazon Comprehend, Amazon Forecast, Amazon Rekognition, Amazon Textract, Amazon Kendra, and Amazon Translate.

| AI Features |
| --- |
| **AI techniques/approaches:**<br>AWS pre-trained AI services easily integrate with enterprise applications to address common use cases. The services are powered by the same deep learning technology that powers Amazon.com and their ML services. No machine learning experience is required of the client. |
| **What functionality does AI enable:** ML; Use ready-made, purpose-built AI services including:<br>• Analysis of images and videos<br>• Text and Data Extraction<br>• Automated speech recognition<br>• Speech translation |
| **Software Offerings (AI-enabled features):**<br>AI pre-trained services integrated with several vendor products (e.g., Hyland Alfresco, Iron Mountain InSight) include:<br>1. *Amazon Textract* is a machine learning (ML) service that automatically extracts text and data from scanned documents, forms and tables. Combined with an Intelligent information processing service, you can classify data from a mass of ingested data.<br>2. *Amazon Comprehend* is a natural language processing (NLP) service that uses machine learning to find insights and relationships in text without machine learning experience—it examines unstructured data such as customer emails, support tickets, product reviews, social media, and even advertising copy for insight into customer sentiment.<br>3. *Amazon Transcribe* uses automatic speech recognition (ASR) to convert speech to text.<br>4. *Amazon Rekognition* uses deep learning technology (DL) and AWS image recognition to extract textual context from images and video and store that information as a document with metadata. |
| **How are functions enabled?**<br>AWS pre-trained AI services are integrated with other applications to address common use cases. Some vendors integrate AWS AI services into their products, and the client purchases that vendor's products. However, other AWS AI services clients seek solutions that work with their own systems. For example, a case study posted to the AWS website explains how Anthem (US health insurance provider) uses AWS Textract to index and classify documents: |

# AI Enabled Services and Software

After a medical provider submits documents to Anthem's provider portal, the documents enter a data store and are sent to a digital-processing function powered by Amazon Textract. Inside this function, Amazon Textract uses optical character recognition (OCR) to extract data and uses machine learning to automatically index and classify each document. The processed claim is then made available to Anthem stakeholders.[1]

**What descriptions of algorithms are provided?**
Detailed descriptions of models were not located. Information videos, such as the one about Amazon Rekognition and ML describe how to work with the model but not the inner workings of the model itself. **https://www.youtube.com/watch?v=fnFpE6UEd_I**

In addition, key features of each service are described. For example, label detection and image properties of Rekognition is described as detecting brightness, sharpness, and contrast so that metadata can be generated for image libraries for search and filtering as well as to identify the quality of the images.[2]

**Third party services supporting AI?**
None.

**Opportunity for User input? If rule-based, can it be customized?**
Users can customize models. For example, when using *Amazon Rekognition* (computer vision service), users can automate machine learning (AutoML) to train models to detect objects such as brand logos with as few as 10 images.

**What information about the AI models in use is provided/available to the user?**
Information is available on the website through product pages, videos, and additional resources on the Customer Enablement page.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
N/A

**Assessment:** Amazon is a noted leader in AI & ML with enough capabilities to offer AI and ML as a service to multiple vendors, including Hyland and Iron Mountain. In addition, use cases found on the customer resources page illustrate the benefit of employing AWS AI services with other processes.

**Recommendations:** The AA01 study group *recommends* learning more about AWS AI Services through an interview with a representative who can verify our findings and provide additional information on the AI-features.

---

[1] https://aws.amazon.com/textract/customers/?pg=ln&sec=hs#Anthem
[2] https://aws.amazon.com/rekognition/image-features/?nc=sn&loc=3&dn=2

# AI Enabled Services and Software

**Name of Product:**               Castlepoint Systems
**Name of Product / Service:**    Castlepoint
**Self-Described Service:**       Information Governance, Risk and Compliance Solution
**Website Address:**             https://www.castlepoint.systems/solutions/
**Gartner CSP Magic Quadrant:**  Yes / No
**G2 Grid for ECM:**              Yes / No

---

**Introduction:** "Castlepoint is a single solution to manage all of the information in an organization. It registers every record in every business system, and uses Artificial Intelligence to classify it against rules and regulations (including Records Authorities), and apply appropriate lifecycle controls."[1]

Castlepoint can help organizations meet the privacy and cybersecurity commitments providing comprehensive information asset management, security management, tracking and reporting on all actions on classified, sensitive, high-risk, and operational records. Records and information management features are powered by AI, enabling accurate capture, fast retrieval, compliance with obligations and information risk management.

| AI Features |
| --- |
| **AI techniques/approaches:** <br> • Records Management Services: 1) Castlepoint reads and registers every item in every system and uses AI to classify it and determine appropriate lifecycle controls. 2) Applies records controls for the whole lifecycle without requiring the user to train the AI. <br> • AI is also used for Discovery, Audit and Compliance, and Cyber Security. |
| **What functionality does AI enable:** <br> For *Classification* and *sentencing* to manage records in other systems, such as SharePoint, Teams and Exchange. <br> To *register, classify, index, and sentence* records in share drives as well as SharePoint. <br> For *search and retrieval* to find key terms related to a specific topic across multiple large business systems (one case study resulted in 99.9% accuracy). The Castlepoint team works with the business/department to develop a taxonomy set, using ontology and regular expression to define terms in scope. |
| **Software Offerings (AI-enabled features):** <br> See previous information. |
| **How are functions enabled?** <br> Through algorithms provided by Castlepoint. |
| **What descriptions of algorithms are provided?** <br> No information located. |

---

1. [1] https://www.cmtedd.act.gov.au/economic-development/innovation-industry-investment/defence/companies/defence/castlepoint-systems#:~:text=Castlepoint%20is%20a%20single%20solution,and%20apply%20appropriate%20lifecycle%20controls.

# AI Enabled Services and Software

| |
|---|
| **Third party services supporting AI?**<br>No, they are provided by Castlepoint. |
| **Opportunity for User input? If rule-based, can it be customized?**<br>The website indicates the vendor and client will work together to set up the solution. However, no information on training and managing AI models. |
| **What information about the AI models in use is provided/available to the user?**<br>None could be found on the website. |
| **Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**<br>N/A |

**Assessment:** Castlepoint provides itself in offering true in-place, continuum management of every record in every business system. It registers every record and uses Artificial Intelligence to classify it against rules and regulations (including Records Authorities) and apply appropriate lifecycle controls.

**Recommendations:** The AA01 study group *recommends* learning more about Castlepoint's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

**Name of Vendor:**             Collabware
**Name of Product or Service:**    Collabspace Archive, Collabspace Discovery, Collabspace Continuum,
                                      Collabmail, Collabware CLM
**Website Address:**             https://collabware.com/
**Gartner CSP Magic Quadrant:**  Yes / No
**G2 Grid for ECM:**            Yes / No

**Introduction:** "Providing Intelligent ECM solutions and services to free organizations from information chaos. Content enrichment harnesses AI to provide richer metadata for auto-categorizing, searching, and summarizing structured and unstructured data."

## AI Features

**AI techniques/approaches:**
- Named entity recognition uses machine learning to automatically identify entities in unstructured data (i.e., scanned pdfs, images, videos);
- Object detection uses a machine learning model to identify named objects within images and apply
- keywords as descriptive metadata; and Sentiment analysis analyzes text to determine the writer's tone and whether this is positive or negative.

**What functionality does AI enable:**
Functionality includes:
- Automated indexing
- Machine learning
  - Keyword extraction
  - Entity extraction (e.g., people, places, etc.)
  - OCR text extraction
  - Audio and video transcription
  - Object detection in images
  - Sentiment analysis
  - Categorization/auto-classification to a file plan
  - Auto-populate metadata values

**Software Offerings (AI-enabled features):**
Collabware sells five products designed to support MS SharePoint:
1. Collabspace ARCHIVE: A data protection application that enables cross-system content access, auditing, recovery, and search. Content connectors automatically stream and index files into a data lake to achieve comprehensive visibility, multi-level permissioned access and effective collaboration.
2. Collabspace DISCOVERY: An enterprise-wide search application. Content connectors stream and auto index files into a data lake to achieve comprehensive visibility, multi-level permissioned access and effective collaboration.
3. Collabspace CONTINUUM: Connects all electronic, physical, and email sources for unified cross-platform visibility and retention. Users work in the systems they're familiar with and records management compliance and backup are automated in the background.
4. Collabmail: Integrates MS Outlook with SharePoint (i.e., user can see SP sites in Outlook).

# AI Enabled Services and Software

**5.** Collabspace CLM: [compliance]; Provides enterprise disposition management.

**How are functions enabled?**

There are 3 functions: named entity recognition, object detection, and sentiment analysis.

1. *Named entity recognition* is an information extraction process that uses machine learning to automatically identify entities in unstructured data, such as images, videos, and scanned pdfs.
2. Using a machine learning model to identify named objects within images, *object detection* automatically provides image keywords to categorize and sort through heaps of image files based on the objects within. The image keywords can also be used to identify PII for redaction before disclosure to a third party.
3. The Collabspace *sentiment analysis* feature will produce a single value between -1 (Negative) and 1 (Positive) that specifies the text's attitude.

**What descriptions of algorithms are provided?**

Collabspace utilizes AI and machine learning (ML) techniques to process data, index continually, and search its proprietary data lake survive, which runs on Microsoft's Azure Cloud (each customer onboarded as a separate tenant).

Discover features: All of the data (information) in the data lake is analyzed and processed via the ML capabilities. New values are extracted from that content and, once added, also stored in the data lake to provide a rich body of active and automatically updated content to mine.

**Third party services supporting AI?**

Most of the work is customized.

- Keyword extraction is based on YAKE
  http://yake.inesctec.pt/

- Named entity recognition is based on spaCy
  https://spacy.io/usage/linguistic-features

- Object detection for images and videos is based on YOLO
  See https://towardsdatascience.com/whats-new-in-yolov4-323364bb3ad3

- Sentiment analysis is based on WordNet
  https://wordnet.princeton.edu/

**Opportunity for User input? If rule-based, can it be customized?**

None. In the future, the plan is to allow customers to leverage their own trained models and add rules to entity recognition.

**What information about the AI models in use is provided/available to the user?**

Users can view the AI features when they review the details of the records in Collabspace, but the data is not downloadable in any way or saved on Collabware servers. This means that the information is technically provided to the users, but not in a way that could be leveraged for analytics—the ultimate goal.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**

In the future, the users will be able to tweak the rules for entity extraction.

# AI Enabled Services and Software

**Assessment:** Collabware, through its three electronic content management (ECM) solutions, provides an innovative approach to information governance challenges in general and RM practices in particular. The solutions are: Collabware CLM: DoD 5015.2-certified SharePoint records management. Collabmail: Email management automation with Outlook and SharePoint integration. Collabspace: Streams all content sources into a data lake to achieve cross-system discovery, records compliance, disposition, analytics, and insights. Given its overt advert mentions of ML and AI use in systems where others are also using ML and AI, Collabware should receive further review and demonstration.

**Recommendation:** The AA01 study group *recommends* learning more about Collabware's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

**Name of Product:**                 Feith Systems

**Name of Product / Service:**      The Feith Platform (enterprise content management, records
                                             management, case management, task management, and physical records)

**Self-Described Service:**          Security-First Records Platform

**Website Address:**                    https://www.feith.com/

**Gartner CSP Magic Quadrant:**    Yes / No

**G2 Grid for ECM:**                  Yes / No

---

**Introduction:** Feith is an auto-categorizer using rules, workflows and AI text analysis.[1] It learns to make decisions based on a combination of word matching, phrase matching, metadata matching, and knowledge accumulated in a training process for text matching. This learning and decision-making functionality makes Feith well-suited to conquering unstructured data repositories, like email, network drives, SharePoint and more.

| AI Features |
|---|
| **AI techniques/approaches:**<br>Advanced statistical modeling to read and categorize documents and their associated metadata. |
| **What functionality does AI enable:**<br>Feith learns to make decisions based on a combination of word matching, phrase matching, metadata matching, and knowledge accumulated in a training process for text matching. This learning and decision making functionality makes Feith well-suited to conquering unstructured data repositories, like email, network drives, SharePoint and more. |
| **Software Offerings (AI-enabled features):**<br>Auto-categorization for records management and information governance. |
| **How are functions enabled?**<br>The following steps are taken: 1) Set up categorization rules; 2) capture documents into the archive; categorize on metadata content; compare documents with the training set; apply retention and classifications. |
| **What descriptions of algorithms are provided?**<br>None. |
| **Third party services supporting AI?**<br>No. This is a Feith product. But it works to categorize structured and unstructured data in other systems, such as email, network drives, Sharepoint and more. |
| **Opportunity for User input? If rule-based, can it be customized?**<br>Yes, auto-categorization rules are based on the organization's industrial, organizational, or governmental requirements. |
| **What information about the AI models in use is provided/available to the user?**<br>None. |

---

[1] https://www.feith.com/rmaiq/auto-categorizer/

# AI Enabled Services and Software

> **Alternatively, if a rules-based system, how much can the user 'tweak' the rules? ?**
> Feith uses rules, workflows, and AI text analysis for both electronic and scanned paper records based on the organization's own industrial, organizational, or governmental requirements.

**Assessment:** The Feith Platform is first a records management platform. Feith software uses AI and ML to conduct electronic content management for electronic and scanned documents found in email, network drives, SharePoint, and more. The software performs extensive actions based on the records lifecycle. Of interest also to our study is the fact that the solution uses machine learning and compares training sets.

**Recommendations:** The AA01 study group *recommends* learning more about Feith's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

**Name of Vendor:**                   Gimmal
**Name of Product/Service:**     Gimmal Records, Gimmal Discover, Gimmal Physical and more
**Self-Described Service:**         Information Governance
**Website Address:**               https://www.gimmal.com/
**Gartner CSP Magic Quadrant:** Yes / No
**G2 Grid for ECM:**              Yes / No

**Introduction:** Gimmal launched in 2002 to provide records management and information governance software. Gimmal Records manages the entire lifecycle of records and information in Microsoft 365, SharePoint, Exchange, Box, Documentum, File Shares, and more. Gimmal Discover can locate, classify, and manage data to migrate privacy risks and protect sensitive information. Gimmal Physical extends governance (including application of retention policies) of physical records and information located in-house or offsite. Gimmal has partnered with Microsoft over a period of 18 years to develop solutions to enhance records management capabilities in Microsoft 365 (including SharePoint and Teams).

## AI Features

**AI techniques/approaches:**

Gimmal's Discover uses a rule-based "text contains" approach to examine text for keywords, phrases, patterns, or regular expressions to find social security numbers, credit cards, and other predictable patterns.

Notably, text matching does not work for some types of information, including street addresses.

In 2021, Gimmal introduced their first static natural language processing (NLP) model designed to identify name and/or address information. The reviewer triggers a workflow decision where results include a threshold percentage for the degree of certainty. The reviewer can set the threshold to only accept results above a threshold (e.g., 80 percent).

The company plans to develop machine-learning capabilities to streamline eDiscovery.

**What functionality does AI enable:**

Rules-based text matching methods in place; natural language processing methods in development.

**Software Offerings (AI-enabled features):**
- Gimmal's Discover module appears to support auto-classification.
- Specifically for Microsoft 365, Gimmal enhances the M365 E5, license with a file plan builder, trainable classifiers, automated retention workflows, multiple reviews, and migration (for OT Content Server).

**How are functions enabled?**

Auto-classification appears to be supported through text matching

**What descriptions of algorithms are provided?**

No information located.

**Third party services supporting AI?**

Gimmal references SharePoint Syntex as a product capable of no-code machine learning where users teach the artificial intelligence model to read content and recognize critical pieces of information,

which is then applied to metadata columns. According to Gimmal, Syntex can be trained to recognize a record and extract and tag metadata including document type, PII, and date created.  The Gimmal Records product apparently leverages this metadata to create event-based, rule-based, and date-based triggers.

**Opportunity for User input? If rule-based, can it be customized?**
Users can set rules.

**What information about the AI models in use is provided/available to the user?**
Users know what parameters they enter.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
The users can set rules.

**Assessment:**  Because Gimmal was launched in 2002 to provide records management and information governance software, it is aware of the features desired by records and information managers related to retention and disposition. Its physical records component makes it attractive to those dealing with hybrid records management challenges. Although many of its solutions are geared to Microsoft products, its ability to locate content for regulatory compliance, litigation, or investigations held in a variety of corporate data sources including local workstations, file shares, PST files, Box, Google Workspace, and more means it is an option that would be of interest to records and information professionals and should be explored further.

**Recommendation:** The AA01 study group *recommends* learning more about Gimmal's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

**Name of Vendor:**                Hyland
**Name of Product / Service:**    Alfresco
**Self-Described Service:**        Content Services Platform
**Website Address:**             https://www.alfresco.com/
**Gartner CSP Magic Quadrant:**  Yes[1] / No
**G2 Grid for ECM:**              Yes / No

**Introduction:** Hyland is a recognized leader by Gartner and Forrester for Content Services Platforms. The Alfresco product is labeled as a "niche" product on the G2 Grid for Enterprise Content Management. Alfresco offers two categories of solutions of special interest to records and information managers: 1) its *Content Services* (Enterprise Content Management software) which includes Document Management; Business Intelligence, Analytics, and Insights; and *Alfresco Intelligent Services* which leverages AI and machine learning from Amazon Web Services (AWS) and 2) its Governance Services which includes Records Management; Information Lifecycle Management; Regulatory Compliance, and eDiscovery and Legal Holds.

## AI Features

**AI techniques/approaches:**
ML, NLP, DL through integration with AMAZON products/tools.

**What functionality does AI enable:**
What functionality does AI enable: Machine Learning (ML) for text extraction; Natural Language Processing (NLP) for insights into relationships in unstructured text documents; and Deep Learning to extract textual context from images and videos.

**Software Offerings (AI-enabled features):**
Both Alfresco's Content Services and its Intelligence Services employ Amazon AI technology: Textract (ML), Comprehend (NLP), and Rekognition (DL).
- Text and data extraction
- Insights and relationships in text
- Image recognition
- Text extraction from images and videos

**How are functions enabled?**
1. Amazon Textract is a machine learning (ML) service that automatically extracts text and data from scanned documents. Combined with Alfresco's Intelligent information processing service, you can classify data from a mass of ingested data.
2. Amazon Comprehend is a natural language processing (NLP) service that uses machine learning to find insights and relationships in text without machine learning experience—it examines unstructured data such as customer emails, support tickets, product reviews, social media, and even advertising copy for insight into customer sentiment.
3. Amazon Rekognition uses deep learning technology (DL) and AWS image recognition to extract textual context from images and video and store that information in Alfresco Content Services as a document with metadata.

---

[1]Hyland's 3 offerings: Alfresco, Nuxeo, and OnBase were included on the Gartner Magic Quadrant for Content Service Platforms as one.

# AI Enabled Services and Software

**What descriptions of algorithms are provided?**
Alfresco uses AMAZON services, so they designed a process and then use ML (for example) to extract data from documents that are uploaded into the system. It can then populate the database with the extracted data (ex. from driver's license for ID or financial statements for financial verification for home loan).

**Third party services supporting AI?**
Amazon Web Services

**Opportunity for User input? If rule-based, can it be customized?**
The developer, data scientist and expert practitioner can customize the out-of-the-box solutions or build their own.

**What information about the AI models in use is provided/available to the user?**
AWS Partners share videos, training and certification center, customer resource center, hands-on tutorials

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
N/A

**Assessment:** Alfresco, part of the Hyland family, is a commercial, open source software company, founded in 2005 with the goal of providing the right information to the eight people exactly when they need it. Alfresco Intelligence Services gives you a scalable way to automatically enrich content and gain valuable insights by leveraging AI and machine learning from Amazon Web Services (AWS), like Amazon Textract, Amazon Comprehend, and Amazon Rekognition. AWS are becoming ubiquitous in the AI, ML, and SaaS market, and this is a good example of a product that integrates AMAZON AI services.

According to Gartner, there is a significant overlap between Alfresco and Nuxeo in terms of both capabilities and target clients. The client should carefully compare each product before making a decision to employ either offering.

**Recommendations:** The AA01 study group *recommends* learning more about Hyland/Alfresco's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

The study group *further recommends* learning more about the AI capabilities integrated into Alfresco from Amazon.

# AI Enabled Services and Software

**Name of Vendor:**           Hyland
**Name of Product/Service:**     Nuxeo
**Self-Described Service:**       Nuxeo Platform for Digital Asset Management (DAM)
**Website Address:**             https://www.hyland.com/en/platform/product-suite/nuxeo
**Gartner CSP Magic Quadrant:**  Yes[1]/No
**G2 Grid for ECM:**            Yes/No

---

**Introduction:**  Hyland is a recognized leader by Gartner and Forrester for Content Services Platforms. The Nuxeo  product is labeled as a "niche" product on the G2 Grid for Enterprise Content Management. Nuxeo is a content services platform for document management, digital asset management, and claims management.  While Nuxeo is branded a platform for digital asset management, it claims to also be an enterprise document management system with governance and retention capabilities.

## AI Features

**AI techniques/approaches:**
Machine Learning and Intelligent Automation
From their website: "Artificial intelligence to predict, classify and enrich content." This is accomplished using *Nuxeo Insight.*

**What functionality does AI enable:**
From their website: "Classify, predict, and enrich documents, rich media, and other content, with no coding or machine learning expertise." Features enabled include:
- Records classification
- Intelligent document capture
- Intelligent data extraction
- Core system integration
- Data extraction
- Data acquisition
- Data validation

**Software Offerings (AI-enabled features):**
- *Nuxeo Insight* uses artificial intelligence for classification and predictive analytics.
- AI is employed to automatically tag images and link related content, improving transparency of assets across the organization and reducing manual labor.

**How are functions enabled?**
- Automated data capture
- Intelligent data extraction

**What descriptions of algorithms are provided?**
No information provided.

**Third party services supporting AI?**
The Nuxeo Platform does not appear to rely on third party solutions.

---

[1]Hyland's 3 offerings: Alfresco, Nuxeo, and OnBase were included on the Gartner Magic Quadrant for Content Service Platforms as one.

# AI Enabled Services and Software

**Opportunity for User input? If rule-based, can it be customized?**
Yes, according to a product brochure available on the website, low-code customization enables organizations to customize their solution and flexible content and data models provide configurable workflows and object models tailored to the organization's needs.[2]

**What information about the AI models in use is provided/available to the user?**
Some information is available on the website[3] including the following steps:
- Train your model by selecting various values and data types.
- Fine tune your data by applying different parameters for each prediction.
- Compare performance to select the appropriate content and data for successful model training.
- Act by performing any necessary renditions or transformations.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
Yes, see previous remarks.

**Assessment:** According to Gartner,[4] there is significant overlap between Alfresco and Nuxeo in terms of both capabilities and target clients. Nuxeo is a newer addition to the Hyland family which uses proprietary AI technology, *Nuxeo Insight*. Although Nuxeo offers enterprise content management, document management, and case management solutions, Nuxeo's strength appears to be in its digital asset management capabilities.

**Recommendation:** The AA01 study group *recommends* learning more about Hyland Nuxeo products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

Because of the perceived overlap between Hyland products and the additional AI features being added, such as *Hyland Brainwa*re for intelligent data capture, it would be advisable to solicit information and a demonstration of all products/services that can be used for lifecycle content and records management.

---

[2] https://www.hyland.com/en/platform/product-suite/nuxeo
[3] https://www.nuxeo.com/content-services-platform/ai/
[4] https://www.hyland.com/en/explore/gartner-magic-quadrant-for-content-services-platforms

# AI Enabled Services and Software

**Name of Product:** Hyland
**Name of Product / Service:** Hyland OnBase
**Self-Described Service:** Enterprise Information Platform
**Website Address:** https://www.hyland.com/en/onbase
**Gartner CSP Magic Quadrant:** Yes / No
**G2 Grid for ECM:** Yes / No

**Introduction:** Hyland is a Leader in this Magic Quadrant. It currently markets three products — OnBase, Alfresco and Nuxeo — as CSPs. Hyland's product portfolio addresses all the content services functional requirements identified by Gartner. OnBase is Hyland's flagship product focused on the midmarket and enterprise space. It is a single enterprise information platform designed to manage content, processes and cases. OnBase centralizes important business content in one secure location, and then delivers relevant information when and where needed. Benefits of this product are increased productivity, excellent customer service, and reduced risk across the entire enterprise.

| AI Features |
| --- |
| **AI techniques/approaches:** <br> Analytics, machine learning and intelligent automation |
| **What functionality does AI enable:** <br>    a.    Records classification <br>    b.    Intelligent document capture <br>    c.    Intelligent data extraction <br>    d.    Core system integration <br>    e.    Data extraction <br>    f.    Data acquisition <br>    g.    Data validation |
| **Software Offerings (AI-enabled features):** <br> Brainware for capture. |
| **How are functions enabled?** <br> Machine learning capabilities identify documents, text, and images without the need for templates. Even different file types in different languages across departments become accessible, providing key metrics and performance indicators. See the Brainware evaluation for additional information. |
| **What descriptions of algorithms are provided?** <br> See the Brainware evaluation for some information on algorithms. |
| **Third party services supporting AI?** <br> All are Hyland products, although two popular services, Alfresco and Nuxeo, were recently acquired. |
| **Opportunity for User input? If rule-based, can it be customized?** <br> Intelligent automation allows line of business users to configure applications in the content management platform. |

# AI Enabled Services and Software

> **What information about the AI models in use is provided/available to the user?**
> See the Brainware evaluation for information related to AI and capture.

> **Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
> Technologies like workflow automation, business logic, machine learning and analytics, are combined to drive digital transformation. No information on 'tweaking' the rules was discovered.

**Note on Integration/Compliance:**
In 2019 Iron Mountain and Hyland announced the integration of Iron Mountain® Policy Center solution with Hyland's OnBase to enable new capabilities to automate retention policy application for records and information stored in Hyland's OnBase enterprise information platform.[1]

**Assessment:**
Hyland's OnBase is an enterprise information platform designed to manage content, processes, and cases. OnBase centralizes important business content in one secure location. Automation, business logic and machine learning are employed to transform processes, reduce errors, and make better decisions. Records retention rules can be applied through integration with the Iron Mountain® Policy Center solution.

**Recommendations:**
The AA01 study group *recommends* learning more about Hyland's OnBase through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

Because of the perceived overlap between Hyland products and the additional AI features being added, such as *Hyland Brainwa*re for intelligent data capture, it would be advisable to solicit information and a demonstration of all products/services that can be used for lifecycle content and records management.

In addition, a demonstration from Iron Mountain® Policy Center either in conjunction with OnBase or as a part of an Iron Mountain product demonstration is warranted.

---

[1] https://www.ironmountain.com/about-us/newsroom/press-releases/2019/september/iron-mountain-and-hyland-partner-to-automate-document-retention-and-improve-compliance

# AI Enabled Services and Software

**Name of Vendor:**                      Hyland
**Name of Product/Service:**       The Hyland Product Suite
**Self-Described Service:**         Content Services Platform and Brainware Intelligent Capture Platform
**Website Address:**               https://www.hyland.com/en/innovation/artificial-intelligence#
**Gartner CSP Magic Quadrant:**  Yes / No
**G2 Grid for ECM:**              Yes / No

**Introduction:**   The Hyland product suite contains software and solutions, some native to Hyland and others gained through acquisition. Careful consideration of the organization's needs, Hyland's offerings, and similar solutions from other vendors is necessary to understand if any of these products would benefit the organization. Three of the products are considered Leaders in the 2021 Gartner Magic Quadrant for Content Services Platforms: Alfresco, Nuxeo, and OnBase.  Other Hyland products are ShareBase, Brainware, Acuo, NilRead, PACSgear, and Enterprise Search.

AI capabilities like analytics, machine learning, and intelligent automation are a native part of Hyland's Content Services platform. Brainware Intelligent capture employs machine learning capabilities to identify documents, text, and images without the need for templates.

| AI Features |
| --- |
| **AI techniques/approaches:**<br>Brainware Intelligent capture is a solution that can extract data from different information sources—from forms with consistent data points to email messages to a handwritten note. |
| **What functionality does AI enable:**<br>• Classification<br>• Extraction<br>• Verification<br>• Integration and Delivery |
| **Software Offerings (AI-enabled features):**<br>Brainware for capture. |
| **How are functions enabled?**<br>Brainware intelligent capture occurs through pattern recognition by identifying where clusters of tabular data like and focus on those for extraction. Integration tools allow Brainware to connect to core business systems for updating and data validation. It learns from historical and real-time data corrections and manual data to improve its ability to capture information accurately. |
| **What descriptions of algorithms are provided?**<br>A neural network of 13 different engines and algorithms:<br>• Improve image quality prior to document classification and data extraction<br>• Create reliable machine learning to continuously improve and accelerate document classification<br>• Enable human-like pattern recognition to boost data extraction accuracy |
| **Third party services supporting AI?**<br>Brainware is a Hyland product. |

# AI Enabled Services and Software

| |
|---|
| **Opportunity for User input? If rule-based, can it be customized?**<br>The product learns on its own through a neural network. User interaction with the models was addressed in the materials reviewed. |
| **What information about the AI models in use is provided/available to the user?**<br>Brainware is based on a neural network of 13 different engines and algorithms. But details on the algorithms could not be located. |
| **Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**<br>N/A |

**Assessment:**  Hyland has a suite of solutions for management of content from creation. The AI technologies offered by the Brainware Platform provide functions related to the early stage of the content/information lifecycle, especially classification, data extraction, and verification. However, other products such as Alfresco, Nuxeo, and OnBase provide features that replicate or complement those functions, such as Nuxeo's ability to classify, predict and enrich documents. It is because of Hyland's reputation within the content services platform and enterprise content management space that each deserves a more detailed analysis.

**Recommendation:**  The AA01 study group *recommends* learning more about Hyland's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

Because of the perceived overlap between Hyland products and the additional AI features being added, such as *Hyland Brainwa*re for intelligent data capture, it would be advisable to solicit information and a demonstration of all products/services that can be used for lifecycle content and records management.

# AI Enabled Services and Software

**Name of Vendor:**            IBM
**Name of Product/Service:**     IBM Cloud Pack for Business Automation
**Self-Described Service:**       IBM Content Services Multi-tenant SaaS on AWS
**Website Address:**          https://www.ibm.com/enterprise-content-management
**Gartner CSP Magic Quadrant:**   Yes / No
**G2 Grid for ECM:**           Yes / No

**Introduction:** IBM is recognized as a niche player on Gartner's 2021 Magic Quadrant for Content Services Platforms. Its capabilities are supported by FileNet Content Manager (for document management), IBM Enterprise Records (records management), IBM Business Automation Workflow (process automation), IBM Business Automation Insights (reporting), IBM Business Automation Document Processing (content intelligence to support content services use cases) and IBM Datacap (intelligent document capture). IBM FileNet Content Manager is recognized as a Leader on the 2022 G2 Grid for Enterprise Content Management.

| AI Features |
|---|
| **AI techniques/approaches:**<br>IBM provides a number of products within its IBM Cloud Pack for Business Automation. All AI is based on IBM Watson. Especially useful is its Machine Learning capabilities for content classification, data extraction, and image recognition. |
| **What functionality does AI enable**:<br>IBM's Machine Learning capabilities enable content classification, data extraction, and image recognition. |
| **Software Offerings (AI-enabled features)**:<br>*IBM Automatic Document Processing* combines AI with deep learning and low-code tooling to help eliminate manual document processing. Classify and extract information from business documents. It can read, refine, and apply data to downstream applications and workflows. |
| **How are functions enabled**?<br>IBM Watson AI provides the foundation for the AI/ML features. |
| **What descriptions of algorithms are provided**?<br>Pre-trained deep learning models use transfer learning to understand context faster using a smaller number of samples. The user sets up fields for the model to extract. |
| **Third party services supporting AI**?<br>None. |
| **Opportunity for User input? If rule-based, can it be customized**?<br>Users can build and train models for classification and data extraction. Users can review results and correct issues. |
| **What information about the AI models in use is provided/available to the user**?<br>Information on models can be found on the website in the form of text, videos, documentation, and tutorials. |
| **Alternatively, if a rules-based system, how much can the user 'tweak' the rules**?<br>N/A |

# AI Enabled Services and Software

**Assessment:**  IBM's Cloud Pack for Business Automation is a full stack cloud platform with over 170 products and services covering data, containers, AI, IoT, and blockchain. The product that is most closely related to this study is IBM's FileNet Content Manager for document management, which uses AI to deliver deep insights from unstructured content. In addition, clients can build their own machine learning models using Watson's AI. For example, the user can build virtual agents with IBM Watson Assistant or build, run, and manage AI models with IBM Watson Studio.

**Recommendation:** The AA01 study group *recommends* learning more about IBM's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

**Name of Vendor:** Intalio

**Name of Product Service:** Content Services Platform

**Self-described Service**: Digital Transformation Solutions

**Website Address:** https://www.intalio.com/en-us/

**Gartner CSP Magic Quadrant:** Yes / No

**G2 Grid for ECM:** Yes / No

**Introduction:** Intalio is positioned as a visionary on the 2021 Gartner Magic Quadrant as a Content Services Platform for Data Capture, Document Management, Digital Assets Management, and Records Management. It has successfully embedded AI capabilities within its CSP product, providing a rich range of cognitive services that enable content recognition and classification. It aligns these services well with its business solutions to provide usable solutions for its target industries in the areas of Content Services, Process Automation, and Data Governance. Intalio offers efficient automation throughout the entire data lifecycle.

## AI Features

**AI techniques/approaches:**

AI & Machine Learning; Robotic Process Automation (RPA)

**What functionality does AI enable:**
- Data capture
- Content categorization
- Encryption / Decryption (does not appear to be AI enabled)
- Classification of assets
- AI-based tagging to label content

**Software Offerings (AI-enabled features):**
- DATA CAPTURE SOFTWARE - Intelligent Capture powered by AI and Machine Learning technologies. Includes scanning as well as electronic data capture from multiple devices.
- File and Categorization (DOCUMENT MANAGEMENT): Categorize content using both machine learning and rule-based automatic classification. Includes OCR capabilities to convert pictures to text documents to facilitate advanced search.
- Secure Control and Retention (RECORDS MANAGEMENT): Encrypt records using Public Key Infrastructure. Authorized users can also decrypt content.
- DIGITAL ASSETS MANAGEMENT: Computer Vision, Recognition, and Machine Learning techniques / automatic classification of assets, automatic, AI-based tagging to label content with keywords to help extract text, sentiments, objects, and more.

**How are functions enabled?**

Intalio Cognitive Services include:
- *Optical Character Recognition*: Employ AI-powered recognition techniques.
- *Face Recognition and Emotion Detection*: Extract facial attributes, detect emotions, and analyze subject sentiment.
- *Automatic Number Plate Recognition*: Use OCR to read vehicle registration plates in images, and integrate with third-party security, military, and judiciary applications and databases.
- *Object Recognition*: Train the software using machine learning to detect objects within an image or file.

# AI Enabled Services and Software

- *Entity Extraction:* Identify, collect, and classify information (textual extraction, organize using predefined categories, validated data using dictionaries and/or algorithms, implement natural language processing techniques, and more.)

**What descriptions of algorithms are provided?**
None located on the website.

**Third party services supporting AI?**
No third-party AI products were cited.

**Opportunity for User input? If rule-based, can it be customized?**
Of interest to Records Managers, custom retention schedules can be defined with automatic or manual transitions. The Records Management product allows different disposition policies to be set based on time or document condition, and a pre-configured disposition can be launched that requires confirmation prior to permanent deletion.

**What information about the AI models in use is provided/available to the user?**
No information could be located.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
Both AI and rules- and workflows are used.

**Assessment:** Intalio is an international software provider with 30 years of innovation in the field of Content and Process Services. It combines cutting-edge technologies with AI and Machine Learning to introduce advanced Line of Business solutions. Its products fall within 3 main categories: Content Services which includes Document and Records Management; Process Management which includes Digital Signature and Case Management; and Data Governance which includes Data Integration, Processing, and Insights.

**Recommendations:** The AA01 study group *recommends* learning more about Intalio's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

| | |
|---|---|
| **Name of Vendor:** | Iron Mountain |
| **Name of Product/Service:** | Iron Mountain InSight® |
| **Self-Described Service:** | A global information management and content services platform |
| **Website Address:** | https://www.ironmountain.com/services/content-service-platform |
| **Gartner CSP Magic Quadrant:** | Yes / No |
| **G2 Grid for ECM:** | Yes / No |

**Introduction:** Iron Mountain Insight® is a content services platform that combines business process management (BPM) and operational workflow expertise with intelligent document processing (IDP) tools that use technology and techniques such as optical character recognition (OCR), artificial intelligence (AI), machine learning (ML) and robust exception processing.[1] Iron Mountain solutions span across industries to enable scanning and storing, digital mailrooms, account payable, and human resources.

Iron Mountain InSight® is designed to help manage and add structure to an organization's unstructured content, such as documents, audio, videos, and photos. The platform leverages intelligent document processing technology to aggregate and connect physical and digital information, extracting key metadata out of documents and automating document-centric workflows to users get to relevant information faster.

Using AWS AI/ML with Iron Mountain InSight®, Iron Mountain enables the customers to extract the text and data from scanned documents in more than 50 languages and in many file formats including PDF, TIFF, Microsoft Office, etc. The AWS services that power Iron Mountain InSight® solution include Amazon Textract, Amazon Comprehend, Amazon Transcribe, and Amazon Rekognition.

| AI Features |
|---|
| **AI techniques/approaches:**<br>AWS AI/ML |
| **What functionality does AI enable:**<br>AWS AI/ML and Iron Mountain Insight® enable clients to:<br>• Extract Text and Data Extraction from scanned documents<br>• Classify documents using Natural Language Processing (NLP)<br>• Route Documents<br>• Fill Forms<br>• Verify data |
| **Software Offerings (AI-enabled features):**<br>Iron Mountain took part in The National Archives of the UK's "AI for Digital Selection project, which revealed some of the features employed to carry out appraisal and selection of document, emails, datasets, and other information.<br><br>Iron Mountain's Artificial Intelligence (AI) and Machine Learning (ML) solution, InSight®, was used to train the system to recognize candidate records for permeant preservation, detect duplicates for disposition, extract entities, and provide file analysis.[2] |

---

[1] https://www.ironmountain.com/services/workflow-automation
[2] https://www.ironmountain.com/uk/resources/case-studies/t/the-national-archives-case-study?localize=false

# AI Enabled Services and Software

Iron Mountain InSight® can be integrated with Iron Mountain's Policy Center to manage retention and privacy.[3]

**How are functions enabled?**

1. *Amazon Textract* is a machine learning (ML) service that automatically extracts text and data from scanned documents. Combined with an intelligent information processing service, you can classify data from a mass of ingested data.
2. *Amazon Comprehend* is a natural language processing (NLP) service that uses machine learning to find insights and relationships in text without machine learning experience—it examines unstructured data such as customer emails, support tickets, product reviews, social media, and even advertising copy for insight into customer sentiment.
3. *Amazon Transcribe* uses automatic speech recognition (ASR) to convert speech to text.
4. *Amazon Rekognition* uses deep learning technology (DL) and AWS image recognition to extract textual context from images and video and store that information as a document with metadata.

**What descriptions of algorithms are provided?**

Amazon AI components are used. They are:

- *Amazon Textract* is based on deep-learning technology that was developed by Amazon's computer vision scientists to analyze billions of images and videos daily. It can extract words and lines of text for NLP applications. It can automate classification of documents and automatically route those pages to the correct analysis operation for further processing.[4]
- *Amazon Comprehend* is a natural language processing (NLP) service that uses machine learning to find insights and relationships in text. No machine learning experience required.
- *Amazon Transcribe* uses a deep learning process called automatic speech recognition (ASR) to convert speech to text quickly and accurately.
- *Amazon Rekognition* makes it easy to add image and video analysis to applications using deep learning technology that requires no machine learning expertise to use. With Amazon Rekognition, you can identify objects, people, text, scenes, and activities in images and videos, as well as detect any inappropriate content.[5]

**Third party services supporting AI?**

Amazon Web Services

**Opportunity for User input? If rule-based, can it be customized?**

Users are involved in the testing process by providing labelled and unlabeled insets to evaluate InSight® in identifying records relevant to selection criteria.

---

[3] https://www.ironmountain.com/services/policy-center
[4] https://docs.aws.amazon.com/textract/latest/dg/what-is.html
[5] https://aws.amazon.com/free/machine-learning/

# AI Enabled Services and Software

**What information about the AI models in use is provided/available to the user?**
In the Insight® documentation, there doesn't seem to be much detail provided about the models themselves. The type of algorithm is not specified, but some features are called out such as routing documents, filling forms, and checking entered data for errors.

Robotic process automation (RPA) uses intelligent software "bots" to identify and manage repetitive tasks such as copying data between software applications or querying customer databases.

Amazon AI components are integrated, and Amazon provides extensive documentation around its AI services.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
N/A

**Assessment:** The most AI-driven component of the software, robotic process automation (RPA), uses intelligent software 'bots' to identify and manage repetitive tasks such as copying data between software applications or querying customer databases. Valuable third-party AI processes come from Amazon Web Services and enable classification, find insights and relationships, convert speech to text, and add image and video analysis to applications. There is an option to add InSight® to Iron Mountain's Policy Center to manage retention and privacy.

**Recommendations:** The AA01 study group *recommends* learning more about Iron Mountain Insight® (and perhaps Iron Mountain's Policy Center) through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

Because several AI features are enabled by a third-party platform, Amazon Web Services, and the same features are used in at least one other product reviewed, Hyland Alfresco, a demo by Amazon may also be warranted.

# AI Enabled Services and Software

**Name of Vendor:**             Laserfiche
**Name of Product/Service:**    The Laserfiche Platform
**Self-Described Service:**     SaaS Enterprise Content Management Platform
**Website Address:**            https://www.laserfiche.com/solutions/
**Gartner CSP Magic Quadrant:**  Yes / No
**G2 Grid for ECM:**            Yes / No

**Introduction:** Laserfiche's Platform is a cloud-based enterprise content system that provides Intelligent Content Capture, Process Automation, Document and Records Management, and Integrated Solutions. Laserfiche is included on Gartner's 2021 Magic Quadrant for Content Services Platforms as a Visionary and on the 2021 G2 Grid for ECM as a Leader. Content can be brought in directly from cloud drives such as Google Drive and One Drive for Business.

| AI Features |
| --- |

**AI techniques/approaches:**
- Recognizes and extracts data from documents to populate line-of-business applications
- Automatically classify incoming documents and apply appropriate retention schedules to promote consistency.

**What functionality does AI enable:**
- Extraction of data from documents to populate line-of-business applications.
- Classification of documents
- Application of retention schedules

**Software Offerings (AI-enabled features):**
- Intelligent capture
- Machine Learning
- Robotic Process Automation

According to the G2 Grid for ECM, Laserfiche is innovative with cloud, machine learning, and AI to enable organizations to transform into digital businesses.

**How are functions enabled?**
- *Intelligent content capture*—automatically classify documents, apply image enhancements, and extract key information based on configured rules. Text is extracted and indexed alongside the document. Smart Invoice capture is a pre-trained ML model to capture data from invoices. It extracts text that can be used to populate metadata, route invoices and more. Can capture content from dynamic digital forms which allows capture of forms fields and association of files with a particular process. Documents are subject to records management, security, and auditing controls upon import and can be routed for review and approval.
- *Business Process Automation*—for records management. Client can build process flows, enable conditional routing, back-end workflows. RPA bots can enable actions, such as retention requirements. Pre-built workflows are available and configurable.

**What descriptions of algorithms are provided?**
Information about algorithms for checksums, digital signatures and encryption can be found, but AI algorithms used for AI features could not be located.

# AI Enabled Services and Software

**Third party services supporting AI?**
Most appear to be Laserfiche proprietary solutions. However, in Feb. 2022, DoubleYard, Inc., a prominent provider of optical and image character recognition (OCR/ICR) partnered with Laserfiche to integrate AI OCR/ICR into the Laserfiche platform for the recognition of both digital print and handwritten text with a high degree of accuracy.[1]

**Opportunity for User input? If rule-based, can it be customized?**
In some instances, such as workflows.

**What information about the AI models in use is provided/available to the user?**
This information was not located.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
Workflows can be configured by the client.

**Assessment:** Laserfiche is included on Gartner's Magic Quadrant for CSPs of 2021 as a Visionary and as an ECM on the 2021 G2 Grid as a Leader. According to Gartner, "The Laserfiche content services platform is focused on content-centric process automation needs. It includes content services, multichannel capture, information governance, records management, intelligent process automation, collaboration, and business application integrations."[2] However, the ability to recognize, categorize, and link documents to retention schedules is of interest to our study.

**Recommendation:** The AA01 study group *recommends* learning more about Laserfiche products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services at this time.

---

[1] https://www.laserfiche.com/presscenter/doubleyard-technology-alliance-partner/
[2] https://www.gartner.com/doc/reprints?id=1-27P2W24F&ct=211020&st=sb

# AI Enabled Services and Software

**Name of Vendor:**               M-Files
**Name of Product / Service:**      M-Files
**Self-Described Service:**          Metadata-driven Document Management Platform
**Website Address:**               https://www.m-files.com/
**Gartner CSP Magic Quadrant:**   Yes / No
**G2 Grid for ECM:**               Yes / No

**Introduction:** M-Files' AI-powered system connects all documents and information on every business platform and repository in the organization, then analyzes them to place them in context. Records can continue to live in the system they were created or originally stored in—network folders, SharePoint, ECM systems, CRM platforms, file-sharing services and other business platforms and applications. M-Files is considered a Visionary on the 2021 Gartner Magic Quadrant for Content Services Platforms and as a Leader on the 2022 G2 Grid for Enterprise Content Management.

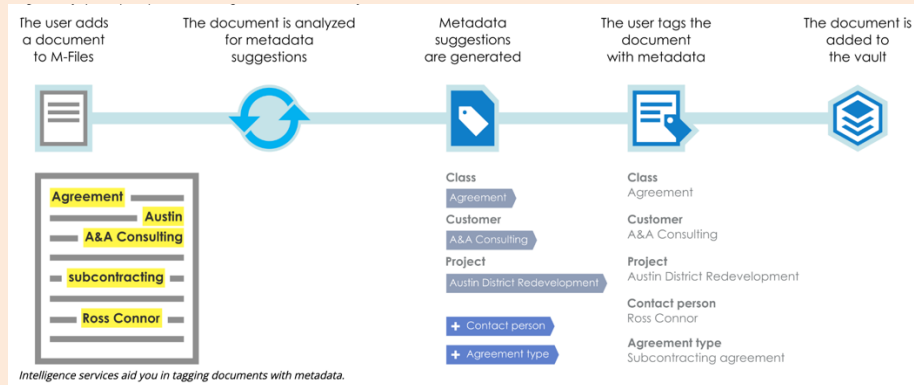| AI Features |
| --- |
| **AI techniques/approaches:**<br>Machine learning; text analytics; natural language understanding; others |
| **What functionality does AI enable:**<br>Identification of personally identifiable information; automatic tagging of files with metadata (classification); discovery of unknown data in legacy archives |
| **Software Offerings (AI-enabled features):**<br>M-Files Intelligent information management; M-Files Smart Classifier; M-Files Discovery<br><br>*M-Files Smart Classifier*: This intelligence service provides classification suggestions for documents added to the vault based on classification models that are automatically and continuously trained and improved in the background by the M-Files Smart Classifier Background Analyzer service. The Smart Classifier consists of two components: the metadata suggestion module and the automatic training module.<br><br>*M-Files Discovery*: M-Files Discovery can be used to automatically classify and categorize documents, as well as find Personally Identifiable Information (PII). It can automatically set metadata, as well as update document permissions and initiate workflows in your M-Files vault. |
| **How are functions enabled?**<br><ul><li>AI is the engine behind the scenes of M-Files driving automation and adding efficiency through functions such as automatic tagging suggestions or discovery of unknown data in legacy archives.</li><li>M-Files for business has AI capabilities for personal identifiable information (PII) data discovery and automatic metadata suggestions.</li></ul> |
| **What descriptions of algorithms are provided?**<br>No information provided; however, the following graphic provides an example of how M-Files' intelligence service can automatically suggest metadata to speed up the process of adding metadata to M-Files objects. |

# AI Enabled Services and Software



The user adds a document to M-Files → The document is analyzed for metadata suggestions → Metadata suggestions are generated → The user tags the document with metadata → The document is added to the vault

*Intelligence services aid you in tagging documents with metadata.*

| | |
|---|---|
| **Third party services supporting AI?** | |
| None specified. | |

**Opportunity for User input? If rule-based, can it be customized?**
AI is used behind the scenes without requiring explicit user interaction/input.

- The *M-Files Classifier* uses classification model that must be trained before using. The Smart Classifier Background Analyzer service will attempt to train a classification model one enough satisfactory documents reside in its vault. The analyzer continues to monitor changes in the vault and will periodically retrain a new classification model. This model can be configured by the user.[1]
- *M-Files Discovery* automatically examines the contents of M-Files Vaults and connected repositories to find and categorize documents based on key phrases and patterns. Documents that contain PII, like credit card numbers and national identification numbers, can be located so that a workflow to handle those documents properly can be initiated.  Paired with the M-Files Classifier, document can be classified regardless of location, such as shared network drives.[2]

**What information about the AI models in use is provided/available to the user?**
None

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
No information available.

**Assessment:** M-Files uses an umbrella term "intelligent information management" (IIM) to refer to their automated document management technologies/services. The services are about automation which is AI powered.  The IIM service features includes detecting PII, monitoring for unusual activity, automating records management activities, finding and deleting unnecessary content and file, and tagging files.  Specific products in the IIM portfolio described as AI powered include the M-Files Smart Classifier and M-Files Discovery.

---

[1] https://m-files.my.site.com/s/article/mfiles-ka-303442
[2] https://catalog.m-files.com/shop/m-files-discovery/

# AI Enabled Services and Software

**Recommendations:** The AA01 study group *recommends* learning more about M-Files' products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

**Name of Vendor:**                  Microsoft
**Name of Product / Service:**     Microsoft 365; SharePoint Syntex; Microsoft Purview
**Self-Described Service:**        Cloud-Powered Productivity Platform[1]
**Website Address:**              https://www.microsoft.com/en-us?ql=3
**Gartner CSP Magic Quadrant:**  <span style="color:red">Yes</span> / No
**G2Grid for ECM:**             Yes / <span style="color:red">No</span>

**Introduction:** Microsoft is considered a Leader on the 2021 Magic Quadrant for Content Services Platforms. M365's most robust platform, E5, supports a range of use cases and is especially strong in improving employee productivity. M365 provides standard content services capabilities, primarily provided by SharePoint, and deep integration into the rest of the Microsoft ecosystem. M365 uses advanced content artificial intelligence (AI) and machine learning to provide intelligent document processing through an M365 service called Microsoft Syntex.[2] Microsoft Purview Data Lifecycle Management provides the tools and capabilities, including pattern-based classifiers to identify sensitive information types and machine learning-based trainable classifiers to identify additional categories of data so that content can be retained if it must be and deleted when it is no longer of value.[3]

With a SharePoint Syntex content center site, you can train AI models to classify and extract data from unstructured documents such as statements of work or contracts. Once you are ready to apply the trained models to your document libraries, you need to start a SharePoint Syntex trial or purchase licenses.

| AI Features |
| --- |
| **AI techniques/approaches**: <br> The capacity for certain defined functionality, ranging from training your own model to the performance of a set task such as language detection or named entity extraction, etc. |
| **What functionality does AI enable:** <br> The feasibility of using some of the above Microsoft services to build a solution for classifying records for retention was tested out as part of TNA's AI for Digital Selection project. The report, Machine Learning for Document Classification,[4] is available for download. The test seems to have involved the use of the applied Cognitive Search service as well as Azure Databricks for training a custom ML model. |
| **Software Offerings (AI-enabled features):** <br> Comprehensive labeling. Microsoft 365 has introduced AI/ML-based automated information governance, allowing organizations to discover and manage information automatically with minimal user interaction. |
| **How are functions enabled?** <br> In terms of the 'out of the box' models you use when using some of these services, there does seem to be information on the versioning of the models - Model Lifecycle of Language service models - Azure Cognitive Services \| Microsoft Docs. And they also provide transparency notes on notes on data privacy |

---

[1] https://support.microsoft.com/en-us/office/what-is-microsoft-365-847caf12-2589-452c-8aca-1c009797678b#:~:text=Microsoft%20365%20is%20our%20cloud,Macs%2C%20tablets%2C%20and%20phones
[2] https://learn.microsoft.com/en-us/microsoft-365/syntex/?view=o365-worldwide
[3] https://learn.microsoft.com/en-us/microsoft-365/compliance/data-lifecycle-management?view=o365-worldwide
[4] https://cdn.nationalarchives.gov.uk/documents/adatis-azure-national-archives.pdf

and security, e.g., Overview of Responsible use of AI - Azure Cognitive Services | Microsoft Docs. Some of these read more like some general information on understanding performance and some disclaimers/lists of things that may affect performance.

**What descriptions of algorithms are provided?**
Some information is provided, although due to the extensive products included in the M365 Platform, it is difficult to cover every solution. For example, a process flow for creating custom classifiers[5] when using Purview illustrates the process from preparing 50-500 positive samples of seed data through creating a classifier and using the model to create a prediction based on the seed data. After approximately 24 hours, the seed data is processed, and test data (up to 10,000 positive and negative samples) is moved to a SharePoint Online folder to be crawled. Once the model evaluates the test items as relevant/not relevant, the results are reviewed by a human who agrees, disagrees, or is unsure (retraining). Once the classifier is ready to be published, it can be used in compliance solutions.

**Third party services supporting AI?**
No.

**Opportunity for User input? If rule-based, can it be customized?**
Users can be involved in training classifiers.

**What information about the AI models in use is provided/available to the user?**
Microsoft produces lots of information about Responsible AI - Responsible AI principles from Microsoft.[6] They provide a range of resources, including guidelines, checklists and technology tools. According to Microsoft, documenting datasets promotes more deliberate reflection and transparency about how the datasets affect machine learning models. You can reach information about Datasheets for Datasets and a template for your own use. [7]

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
N/A

**Assessment:** Microsoft M365 and SharePoint provide content services capabilities deeply integrated into the rest of the Microsoft ecosystem. M365 uses advanced content artificial intelligence (AI) and machine learning to provide intelligent document processing through an M365 service called Microsoft Syntex.[8] Microsoft Purview provides risk and compliance solutions, including management of data governance and records management. Microsoft is one of the few software in this survey that presents detailed explanations of how it uses machine learning in various capacities directly in use by records managers. These three products relate most closely to our study.

**Recommendations:** The AA01 study group *recommends* learning more about Microsoft 365, SharePoint, Syntex, and Purview through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

---

[5] https://learn.microsoft.com/en-us/microsoft-365/compliance/classifier-learn-about?view=o365-worldwide#custom-classifiers
[6] https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6
[7] https://www.microsoft.com/en-us/research/project/datasheets-for-datasets/
[8] https://learn.microsoft.com/en-us/microsoft-365/syntex/?view=o365-worldwide

# AI Enabled Services and Software

**Name of Product:**              Newgen
**Name of Product/Service:**    NewgenONE Digital Transformation Platform

**Self-Described Service:**       Digital Transformation Platform
**Website Address:**              https://newgensoft.com/home-us/
                                https://newgensoft.com/platform/records-management-system/
**Gartner CSP Magic Quadrant:**  Yes / No
**G2 Grid for ECM:**          Yes / No

**Introduction:**

Newgen is a leading provider of a unified digital transformation platform with native process automation, content services, and communication management capabilities. The platform includes a number of solutions including enterprise content management, records management, intelligent document processing, business processing management, digital process automation, and robotic process automation (RPA).

The Records Management Solution (RMS) is an EDRMS (electronic document records management system) that manages both physical and electronic records.  Intelligent Records Classification enables rapid digitization of physical content through scanning and intelligent information extraction. AI and ML is employed to classify records based on content for smarter classifications and metadata generation.

## AI Features

**AI techniques/approaches:**
- Robotic Process Automation (RPA)
- AI & ML powered cognitive services for electronic content management
- AI and ML-based content classification, metadata generation, mobile-based access, and social content management
- AI and ML used in intelligent document processing for Identity document data extraction and redaction
- Robotic Process Automation acts as a bridge between human and system to automate repetitive and mundane tasks.

**What functionality does AI enable:**
- Intelligent Content Classification  (Facilitators: AI & ML) 3 ways to classify unclassified documents: layout / content / object
- Intelligent extraction (including redaction)
- RPA for repetitive and mundane tasks within business workflow
- In addition, their ECM solution uses AI-ML powered cognitive services
- Intelligent Document Processing solution uses AI and. ML for layout- and content-based classification, as well as identity document data extraction and redaction.
- Intelligent Document Processing solution also uses RPA to automate repeatable business processes such as intelligent responses to customers.

**Software Offerings (AI-enabled features):**
- *Intelligent Document Processing features include* AI-based capabilities in imaging, document extraction, redaction, and classification

# AI Enabled Services and Software

- *Automatic classification and extraction* of documents with AI-based bots to achieve straight-through-processing of complex document-based business processes, such as invoice processing, trade finance, etc.
- AI-based bots for *masking and storing critical information* on the images and documents; automating incoming email to allow response and initiation of a business transaction in a third-party system; and to listen and sense customer interactions across various social media channels like Twitter, Facebook, etc., and deliver contextual responses.

**How are functions enabled?**
See information already provided about document, content, and records management solutions. In addition, Newgen offers an AI Cloud[1] that enables clients to visually manage the end-to-end data lifecycle, from preparing and deploying models to monitoring on a single unified platform. Equipped with intelligent automation capabilities, it enables users to perform key machine learning (ML) tasks within a few minutes. Furthermore, it allows citizen and expert data scientists as well as data engineers to efficiently work on AI/ML projects.

**What descriptions of algorithms are provided?**
Most AI-related information is provided for the AI Cloud, which includes the following comments about the option geared to helping the customer make better business decisions and transform the customer experience:
- Use built-in machine learning (ML) and deep learning-based techniques for dimensionality reduction, including singular value decomposition (SVD), principal component analysis (PCA), and restricted Boltzmann machine (RBM)
- Use multiple options to model, including graph, ML, deep learning, and natural language processing
- Use multiple modeling techniques on the same feature engineered data with multi-model experimentation and evaluation

It is not clear what types of algorithms are integrated into the records management and enterprise content management solutions.

**Third party services supporting AI?**
None were discovered.

**Opportunity for User input? If rule-based, can it be customized?**
Content classification includes a model training studio to learn about new content types for classification and metadata extraction.

**What information about the AI models in use is provided/available to the user?**
Basic on website for solutions except the AI Cloud (described previously)

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
Document processing mentions built-in capabilities to manage document-centric workflows.

---

[1] https://newgensoft.com/platform/artificial-intelligence-ai-cloud/

# AI Enabled Services and Software

**Assessment:** Newgen is a Visionary in the Gartner Magic Quadrant and recognized for its strong records management. Newgen focuses on enterprise clients, with the majority of these in the financial services industry. Other significant industries are government, healthcare and insurance. Newgen has been working on enhancing AI capabilities, such as predictive metadata and content recommendations. Its AI Cloud offers a number of opportunities for the client to engage in AI and ML projects.

**Recommendation:** The AA01 study group *recommends* learning more about Newgen's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

**Name of Vendor:**                  OpenText
**Name of Product/Service:**     Opentext Documentum
**Self-Described Service:**         Content Services Platform
**Website Address:**                https://www.opentext.com
**Gartner CSP Magic Quadrant:**  Yes / No
**G2 Grid for ECM:**              Yes / No

**Introduction:** OpenText Documentum is part of Open Text Information Management Content Services Platforms along with OpenText Extended ECM and Open text Core.[1] It facilitates seamless information access distribution and use and integrates into enterprise business applications to enable access from any UI. OpenText is a leader in the 2021 Gartner Magic Quadrant for Content Services Platforms, and Open Text EIM is recognized as a Niche Player on the 2022 G2 Grid for Enterprise Content Management.

| AI Features |
| --- |
| **AI techniques/approaches:** |

- Open-source machine learning and predictive analytics platform.
- Automatic versioning capabilities to manage and track content regardless of format.
- Automatically archive, delete, or revoke access to files that have reached the end of their life cycle.

**What functionality does AI enable:**

- OpenText Magellan "analyzes big content made up of structured and unstructured data stored in enterprise data management and external sources"[2].
- Magellan's Data Science Notebook enables machine learning development, customization and model governance for data scientists and customized dashboards for business analysts and operational users.
- Natural language processing and text-based data mining are employed to easily include unstructured content and facilitate advanced analytics solutions, such as sentiment and emotion analysis tools.

**Software Offerings (AI-enabled features):**

- Leverages machine learning to unlock value within structured, unstructured and IoT data to shape operations and improve efficiency through AI data-driven decision-making.
- Reduces manual processing and enhances enterprise data management through AI-augmented automation
- Exchange and enrich data and content between business applications or systems through a machine learning platform.

**How are functions enabled?**

- Multiple AI and machine-learning approaches are used to enable high-fidelity classification and gain knowledge from unstructured content.
- OpenText Magellan is an AI and analytics platform that offers a library of pre-built AI models, including models for computer vision purposes that enable audio, video and image

---

[1] https://www.opentext.com/assets/documents/en-US/pdf/opentext-po-documentum-en.pdf
[2] https://www.opentext.com/products/magellan-platform

| |
|---|
| classification. |

| |
|---|
| **What descriptions of algorithms are provided?**<br>No description available. |

| |
|---|
| **Third party services supporting AI?**<br>AI components are provided by OpenText. |

| |
|---|
| **Opportunity for User input? If rule-based, can it be customized?**<br>AI & Analytics includes both prebuilt and customizable features. |

| |
|---|
| **What information about the AI models in use is provided/available to the user**?<br>OpenText utilizes predictive modeling to create, validate and save versions of machine learning models based on constantly evolving big data and unstructured content. |

| |
|---|
| **Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**<br>Open Text Magellan is open-source machine learning so it can be "tweeked" by users. |

**Assessment:** Documentum is a high-volume content management platform that is built on a modular foundation. It is cloud-based and integrates with enterprise applications such as SAP and Microsoft, that produce and consume information. The software has authentic AI and ML training capabilities as foundational to its operations. ML can be trained to conduct other archives and records management tasks.

**Recommendation:** The AA01 study group *recommends* learning more about OpenText's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

**Name of Vendor:**          RecordPoint
**Name of Product/Service:**     Records365
**Self-Described Service:**      Data Trust Platform
**Website Address:**         https://www.recordpoint.com/
**Gartner CSP Magic Quadrant:**   Yes/No
**G2 Grid for ECM:**          Yes/No

**Introduction:** Recordpoint offers a Data Trust Platform that allows management of customer data from capture through disposal in one place regardless of where it is located. Records365, the focus of this review, is their automatic and compliant records management solution which allows data and records to be managed in place. Connectors can be provided to thousands of structured and unstructured data sources. Artificial Intelligence and Machine Learning is used to apply rules with a feature known as *classification intelligence*. Records can be disposed of based on classification, and rules can be defined to dispose of data automatically and safely.

| AI Features |
|---|
| **AI techniques/approaches:**<br>RecordPoint uses rules-based classification that examines metadata and files using a rules tree. Machine Learning Classification creates a binary copy of textual images and suggests classification based on context. The model can be trained using 50-100 documents. Once the system is trained, auto-apply can be turned on. This product produces a high degree of confidence. |
| **What functionality does AI enable:**<br><ul><li>Intelligent data classification using custom rules and machine learning to train the classification model.</li><li>AI and ML are used to automatically analyze and classify records and then flag for secure retention, disposition, or holds based on content in files and records.</li><li>Sensitive data signaling using AI to screen and locate sensitive data including Personally Identifiable Information (PII) and Payment and Information (PCI).</li></ul> |
| **Software Offerings (AI-enabled features):**<br><ul><li>Machine Learning Analysis and Classification</li><li>Destruction and Holds Workflows</li><li>Sensitive data signaling</li></ul> |
| **How are functions enabled?**<br>The Data Trust Platform is comprised of records management and data lineage tools that work together throughout the lifecycle of the data and connect to hundreds of essential business systems such as Box, SAP, Google, Salesforce, and more. |
| **What descriptions of algorithms are provided?**<br>A classification model is used. A blog post, "ML 101: How machine learning powers RecordPoint's Classification Intelligence," provided this high-level overview of the model used:<br>1. The model first indexes the document's text content.<br>2. The app then provides a prediction about the type of record the content corresponds to. This is directly correlated to the client's Business Classification Schema (BCS)/File Plan.<br>3. The records manager reviews the policy suggestion and either suggests or corrects it. |

# AI Enabled Services and Software

| |
|---|
| **4.** Once the model is trained, it can automatically classify data. |

| |
|---|
| **Third party services supporting AI?** |
| No third-party AI providers were discovered. |

| |
|---|
| **Opportunity for User input? If rule-based, can it be customized?** |
| Yes, rules can be customized. |

| |
|---|
| **What information about the AI models in use is provided/available to the user?** |
| The classification model can be trained using custom rules and machine learning; some information was discovered in a blog post.[1] |

| |
|---|
| **Alternatively, if a rules-based system, how much can the user 'tweak' the rules?** |
| Yes, clients can use their own rules and ML to train classification models. |

**Assessment:** RecordPoint's digital trust platform offers a federated approach to provide automated, intelligent compliance using rules-based and machine learning technology. The product for records management, Records365, supports defensible disposition by using ML for classification and workflows for disposition and holds. It allows for management of email using three approaches: Capstone (select users), department using shared folders; or manage everything.

**Recommendation:** The AA01 study group *recommends* learning more about RecordPoint's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

---

[1] https://www.recordpoint.com/blog/ml-101-how-machine-learning-powers-recordpoints-classification-intelligence

# AI Enabled Services and Software

**Name of Vendor:**                 Relativity
**Name of Product/Service:**     Relativity Assisted Review (incorporating AI); Relativity Analytics
**Self-Described Service:**        Software for legal and compliance
**Website Address:**               https://relativity.com
**Gartner CSP Magic Quadrant:** Yes / No
**G2 Grid for ECM:**                 Yes / No

**Introduction:** Relativity is the market-leader in providing e-discovery services in the US and elsewhere. The company's self-described areas of expertise where their products and services are used include e-Discovery, case strategy, data privacy, early case assessment, investigations, data migration and information governance.

Relativity offers types of AI features and enhancements, including under the label "technology assisted review" (a term well-known to e-discovery lawyers).

| AI Features |
| --- |

**AI techniques/approaches:**

1. For *Relativity's e-discovery related services* using AI methods, it purports to use support vector machine (SVM) methodologies as a core machine learning classifier. According to a Relativity White Paper[1].

   The *active learning Assisted Review workflow* uses a modified LIBLINEAR—a software package that implements a linear SVM—to perform binary classification. LIBLINEAR is a common package used in several open-source machine learning toolkits including GATE, KNIME, Orange, and scikit-learn.

2. *For Relativity Analytics*, the company states that it employs an Analytics index using Latent Semantic Indexing (LSI), for the purpose of grouping documents and identifying conceptual relationships.

3. *Text IQ*, a suite of artificial intelligence workflows to improve document review accuracy; to automate document reviews for privilege, data breach, and privacy; and to use sociolinguistic hypergraph technology to understand context and relationships.

**What functionality does AI enable:**

1. *Relativity's e-discovery related services:* The active learning Assisted Review workflow uses a modified LIBLINEAR—a software package that implements a linear SVM—to perform binary classification. LIBLINEAR is a common package used in several open-source machine learning toolkits including GATE, KNIME, Orange, and scikit-learn.
2. *Relativity Analytics:* Analytics helps reveal the facts of a case by doing the following:
   - Giving users an overview of the document collection through clustering
   - Helping users find similar documents with a right-click
   - Allowing users to build example sets of key issues
   - Running advanced keyword analysis

---

[1] *https://help.relativity.com/9.5/Content/Relativity/Analytics/Analytics.htm*

# AI Enabled Services and Software

**3.** *Text IQ*: Data breach response application to locate personal data. Use Text IQ AI to tag personal data automatically. Transformative labeling to identify columns that contain personal data and then decisions apply across all spreadsheets that contain that data.

**Software Offerings (AI-enabled features):**
Structured Analytics; Conceptual Analytics (index-based); Clustering; Cluster Visualization; Classification; and with Text AI – automatic tagging (labeling). AI-entity extraction and sentiment analysis (used for search).

**How are functions enabled?**
It depends on the feature; for example, entity extraction and sentiment analysis (positive and negative sentiment on sentence-by-sentence basis) enabled with a button from the dashboard will be incorporated in Relativity One soon.

**What descriptions of algorithms are provided?**
- A *recommendation algorithm* is used to determine which documents should be served up for review.
- For classification, *Support Vector Machine (SVM)* technology (a binary supervised machine learning classifier) is used to build a binary classification (called categorization by Relativity) model that codes words as relevant, non-relevant, or both, with low importance placed on those classified as both during the training of the model.

**Third party services supporting AI?**
Relativity products and services do not rely on 3ʳᵈ parties. Relativity can be run as a standalone application but can also be deployed in the cloud using Amazon Web Services or Microsoft Azure.

**Opportunity for User input? If rule-based, can it be customized?**
During the training phase documents coded by reviewers are sent to the model. Reviewers have the option to code a document relevant or non-relevant, or skip a document. Skipped documents are treated as a non-coding decision and are no longer eligible for review by another reviewer.

During the validation phase, the TAR allows for an active learning approach called "elusion." This test estimates the number of relevant documents the model missed. If the elusion is acceptably low, the administrator ends the project; if too high, the administrator resumes review for model improvement.

**What information about the AI models in use is provided/available to the user?**
The white paper "Active Learning in Technology-assisted Review, Relativity's Approach to SVM and the Tech Behind It," January 31, 2018, assists customers in understanding the technology and workflows provided within Relativity. (Updated documentation is found on the Relativity Website. For example, both user and developer documentation for Relativity 1.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
N/A

**Assessment**: TAR has been used in the legal domain for more than a decade. Compliance and legal review

are issues that face organizations with which records managers and information governance professionals are often involved.  The various features within Relativity's offerings, such as labeling, sentiment analysis, classification, and categorization have relevance to our field.

**Recommendation**: The AA01 study group *recommends* learning more about Relativity's products and services through an interview with a representation who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

**Name of Vendor:** Rossum

**Name of Product / Service:** The Rossum Platform

**Self-Described Service:** Document Processing Platform

**Website Address:** https://rossum.ai

**Gartner CSP Magic Quadrant:** Yes / No

**G2 Grid for ECM:** Yes / No

**Introduction:** The Rossum Document Processing Platform was built as a single cloud solution. The software uses AI to conduct pre-processing, data capture, validation, post-processing, content mining and manual review. Data can be extracted from documents stored in multiple file formats, validated by the client, and then exported to another system such as an ERP or accounting system.

## AI Features

**AI techniques/approaches:** AI OCR - Automated data extraction from invoices, purchase orders, packing lists, receipts, or any similar document, including complex table data (built on neural networks). Validation of extracted data integrated with the data capture engine so that the algorithms are improved with every "click and keystroke."

**What functionality does AI enable:**
- In the preprocessing stage, Rossum also manages duplicates, splits documents, and checks auxiliary data and spam.
- In the capture stage, Rossum extracts data from documents (invoice forms for example) with their data capture engine.
- In the post-processing stage, Rossum tracks the journey of each document until it reaches its destination.

**Software Offerings (AI-enabled features):**
Rossum employs AI in the pre-processing, capture, and post-processing phases. It performs AI OCR - Automated data extraction from invoices, purchase orders, packing lists, receipts or any similar document, including complex table data (built on neural networks).

**How are functions enabled?**
The Rossum AI engine (tool that helps build an Artificially intelligent system) runs the Rossum application. Rossum has two types of AI Engine: *Generic AI Engine* and *Dedicated AI Engine*. Rossum has several types of Generic Engines that have learned to process a specific document type. These include Generic Engines learn to recognize fields from many various invoice layouts, languages, and types of content. Generic Engines are default. The **Dedicated AI Engine** learns from unique, business-specific data. As the client reviews, annotates, and validates documents via the Rossum user interface, the engine is automatically being provided with valuable data for learning. Annotations increase data extraction accuracy for the documents.

**What descriptions of algorithms are provided?**
No specific algorithms mentioned aside from the use of AI and ML.

**Third party services supporting AI?**
This appears to be Rossum-created software.

# AI Enabled Services and Software

| |
|---|
| **Opportunity for User input? If rule-based, can it be customized?**<br>Rossum comes preloaded with data fields the platform understands, so you can start using it immediately. You can edit the structure of the fields you want to capture from each queue. It is just a matter of checking and unchecking boxes. |
| **What information about the AI models in use is provided/available to the user?**<br>Information is provided about the AI engine which is described by some as a "group of models."[1] |
| **Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**<br>Yes, some actions are rules-based and geared to the client's needs. |

**Assessment:** This is a very good example of AI OCR. The document process is a major component of records and archives management. The automation features of this software significantly aid in this process using AI, especially when it comes to invoice processing. The software's neural network learns documents and automates content capture and extraction for use in other software. OCR automation is highly useful.

**Recommendation:** The AA01 study group *recommends* learning more about Rossum.ai products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

---

[1] https://www.arabesque.com/2021/09/06/ai-engine-for-the-average-joe/

# AI Enabled Services and Software

**Name of Vendor:**                      SERgroup Holding International GmbH
**Self-Described Service:**         Enterprise Content Management
**Website Address:**               https://www.sergroup.com/en/
**Gartner CSP Magic Quadrant:**  Yes / No
**G2 Grid for ECM:**             Yes / No

**Introduction:**    SER Group was recognized as a Visionary  on the Gartner 2021 Magic Quadrant for Content Service Platforms. Its Doxis4 iECM (on-premises) and Doxis4 Cloud iECM (SaaS) suites support both traditional task-based business processes and case-based business processes, by providing predictive email filing and strong workflow capabilities. The Doxis4 iECM content services platform offers intelligent content automation, content services, content understanding, archiving, and process automation.

| AI Features |
| --- |
| **AI techniques/approaches:**<br>a.  predictive email filing (nudges the users to where an email should be stored in the repository)<br>b.  auto classification<br>c.  metadata augmentation<br>d.  integration of computer vision and NLP for content understanding |
| **What functionality does AI enable:**<br>a.  Filing of email using classification and extraction services<br>b.  Classification<br>c.  Extraction<br>d.  Addition of metadata<br>e.  Content analytics through NLP and computer vision |
| **Software Offerings (AI-enabled features):**<br>The AI capabilities SER added to its suites provide Content Intelligence. Content intelligence is a set of capabilities that enable content classification, metadata augmentation, integration of computer vision and natural language processing to provide content understanding without requiring manual intervention. |
| **How are functions enabled?**<br>According to the website, "Doxis is the first content services platform that has cognitive services at the core of the platform — making all applications running on the Doxis platform more intelligent. These services combine the latest AI technologies with proven methods for classification & extraction, text mining in unstructured documents, and for text analyses."[1] |
| **What descriptions of algorithms are provided?**<br>The solution is called "Doxis AutoML technology." ML methods are integrated into document management & processes to expedite workflows & enhance the quality of results.  See the Doxis4 Cognitive Model in Figure 1. ML methods recognize patterns, find relevant content, extract metadata & help shed light on risks. |

---

[1] https://www.sergroup.com/en/business-platform/content-understanding.html

# AI Enabled Services and Software



**Figure 1:** AI-enabled features of Doxis4 Cognitive Services

**Third party services supporting AI**?
No, this appears to be a SER solution.

**Opportunity for User input? If rule-based, can it be customized**?
The deeply integrated AutoML technology in Doxis enables users to quickly train & deploy new AI models without the need for in-depth machine learning know-how.

**What information about the AI models in use is provided/available to the user**?
Figure 1 shows the various features that are AI-enabled. A visit to the site will allow the user to place a cursor over each feature to learn more. For example, *Extraction* is applied to business documents to extract key terms, extract data from fields and tables, employ rule-based and ML-based actions, employ name entity recognition (NER) and understand correlations. When applied to images, audio & videos, the solution adds metadata, recognizes objects, extracts texts. The *Classification* feature performs layout and content classification, page separation, and document structuring for images/structure. Classification also employs intelligence features such as recognizing sentiment/emotions, recommending experts, and identifying topics.[2] *Language recognition* converts speech-to-text—for example input/audio recordings can be converted to machine-readable text; the feature uses market-leading external tools/APIs. *Natural Language Processing* is used to *understand speech*. It recognizes language, intention, and entities. Interaction examples are making search queries, creating tasks, querying status, and completing tasks.  Intelligent search finds similar content in documents, finds similar text passages, and finds experts. The New Analytics options are significant terms aggregation and auto-clustering of search results. Similarities & recommendations are enabled through ad hoc learning, where the algorithm learns dynamically from any document pools and categorizes documents of a process spontaneously. Recommendation examples include recommending a way to handle a case and adjust workflow steps. Documents/files from similar cases are identified.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules**?
The term rules-based does not appear, but the solutions offered are for specific industries and include

---

[2] https://www.sergroup.com/en/business-platform/content-understanding.html

records retention and compliance—which must be adjusted to meet the needs of the client. It is not clear if this can be done independently or in collaboration with SER Group.

**Assessment:**  While mainly serving clients in Europe, this solution is of interest because of its integration capabilities with SAP, Salesforce, Microsoft, ERP integration and more.  The fact that this is a content services platform that has cognitive services at the core of the platform, along with its integration capabilities, makes it extremely attractive to those looking for one solution to automate its business processes.

**Recommendation:** The AA01 study group *recommends* learning more about SERGroup's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

**Name of Vendor:**            Shinydocs
**Name of Product/Service:**    Cognitive Suite
**Self-Described Service:**     Information Management Software
**Website Address:**          https://shinydocs.com/
**Gartner CSP Magic Quadrant:**  Yes / No
**G2 Grid for ECM:**          Yes / No

---

**Introduction:** Cognitive Suite is a software platform that powers solutions by connecting unstructured and structured information to provide services, including cleanup, search, migration, data, security, compliance, content, records.  According to a document, *3 Steps to Eliminating Shared Drives with Artificial Intelligence*,[1] Shinydocs leverages artificial intelligence and machine learning to organize their documents by finding unstructured information wherever it exists and eliminating redundant, obsolete and trivial files. Shinydocs searches, organizes, and understands unstructured data at scale, providing actionable security, compliance, and business operations insights.

| AI Features |
| --- |
| **AI techniques/approaches:** |
| The Cognitive Suite platform consists of connectors for Exchange Email, OpenText Content Suite and Documentum, Microsoft SharePoint and M365, and possibly other technologies. Shinydocs can create an inventory of records across all repositories and provides various services based on this platform. |
| **What functionality does AI enable:** |
| The functionality is described in terms of eight services:  Content Cleanup and ROT Removal; Discover Search; Migration; Shinydrive; Enterprise Data Quality; Security and Privacy; Content Identification; and Records Identification. |
| **Software Offerings (AI-enabled features):** |
| This is not clear. The Shinydocs white paper, "Initiate a Metadata Analysis by Addressing Your Redundant, Obsolete, and Trivial Data," describes the process in four steps:<br><br>• identify all documents, files, and data to create a complete data ecosystem;<br>• crawl and assess the data to separate records of business value from redundant, outdated, and trivial (ROT) material;<br>• remove ROT material (i.e., delete duplicates and system files, delete or move manage over-retained files); and<br>• tag, search, enrich, and group business records to observe trends, address compliance, and find information faster. |
| **How are functions enabled?** |
| Automated records management including rules-based classification and migration. Using artificial intelligence, the tool extracts entities, identifies document types, data size, data location, and document content. |
| **What descriptions of algorithms are provided?** |
| Information not available. |

---

[1] https://archive.shinydocs.com/wp-content/uploads/2018/03/3-Steps-to-Eliminating-Shared-Drives.pdf

# AI Enabled Services and Software

**Third party services supporting AI?**
There is no mention of third-party provided AI technology. Rather, their Cognitive Suite provides insight into content in third-party repositories (e.g., OpenText, M365) using their AI solution.

**Opportunity for User input? If rule-based, can it be customized?**
Users can employ custom rules to automatically identify records based on their governance structure. The software uses the same rules to classify, tag, and migrate records as they are created.

**What information about the AI models in use is provided/available to the user?**
This is not known.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
Yes, the user has input into rules used for records identification, classification, and migration.

**Assessment:** Shinydocs is not an enterprise content management or records management solution. It has three AI-enabled solutions to facilitate enterprise content management: 1) *Shinydrive* interface which allows users to work with the content in any of their applications, 2) *Migration* which makes an exact copy of all content and metadata to be migrated, and 3) *Cognitive Analysis* which enables rapid analysis of content.

**Recommendation:** The AA01 study group *recommends* learning more about Shinydoc's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# AI Enabled Services and Software

**Name of Vendor:**          Valora Tech

**Name of Product/Service:**     Valora PowerHouse AutoClassification Platform; BlackCat for hosting, metadata management & data visualization

**Self-Described Service:**      Technology provider of automated document data mining and content analytics for the Legal, Records Management & Information Governance markets.

**Website Address:**           valoratech.com

**Gartner CSP Magic Quadrant:** Yes / No

**G2 Grid for ECM:**             Yes / No

**Introduction:** Valora's rich Powerhouse product suite encompasses state-of-the-art content analysis and Autoclassification, incorporating elements of probabilistic systems, Bayesian learning, natural language processing and machine learning (AI). Their software products enable autoclassification, data mining, content analytics, document intake and visualization, rich metadata tagging, they offer consulting services for Project Management and Information Governance.

## AI Features

**AI techniques/approaches:**

Valora primarily uses Guided Machine Learning to perform a Bayesian (rules-based) analysis of the document population. When clients are able to provide specific guidance (e.g., a list of employee names), Valora incorporates that to provide a more parameterized approach, which typically yields stronger accuracy (precision).

**What functionality does AI enable:**

- Rich metadata tagging across hundreds of customizable fields
- Support for thousands of unique Document Types
- ROT identification, sequestration/quarantine, tombstones and removal
- Data extraction, including names, dates, document types, records management fields (e.g., retention period), data privacy fields (e.g., PII), content management fields (e.g., terms), content-specific fields (e.g., sentiment), and customer or employee specific fields (e.g., Emp/Cust Number).
- Records retention identification and tracking
- Migration to other systems, including recommended file location (automatic taxonomies)
- Data privacy detection of personal data, covering PII, PHI and PCI, including AutoRedaction of sensitive content
- Automated security access determination and monitoring
- Data requests and productions for eDiscovery/litigation/FOIA//GDPR DSARs/investigations/search queries
- Identification and management of content lifecycle management, including integration with legal holds
- Business transformation and Robotic Process Automation (RPA)
- Data visualization, data mapping and consolidated views across storage locations

**Software Offerings (AI-enabled features):**

Autoclassification 101 to intelligently automate document processing, analysis, and disposition.

# AI Enabled Services and Software

**How are functions enabled?**
The PowerHouse machine-learning classification engine uses a 5-step methodology to locate, identify, action, and monitor content across multiple data stores. It answers 'where is it' (Crawl & Locate); 'what is it' (Identify & Tag); 'what am I looking at' (Analyze & Understand); 'what do I do with it' (Decide & Action), and 'how often do I update' (Monitor & Audit).

**What descriptions of algorithms are provided?**
The PowerHouse product suite that provides content analysis and autoclassification, incorporates elements of probabilistic systems, Bayesian learning, natural language processing, and machine learning (AI).

The AutoClassification suite of software uses both *pattern-matching algorithms and machine learning* to detect file contents and attributes, and assign contextual attributes (rich metadata) and disposition (rules) for each document or file.

**Third party services supporting AI?**
Valora's uses its own algorithms. Each Valora implementation starts with the baseline functionality and is customized to the client's data set and unique processing/tagging requirements.

**Opportunity for User input? If rule-based, can it be customized?**
Valora encourages client input into the training and tagging models.

**What information about the AI models in use is provided/available to the user?**
Valora provides both a copy of the rules themselves, as well as specific rationales for what each classification call is made. Often they will provide a 2-field pairing between the classification tag and the rationale. Example: ROT value (= Yes or No) and ROT Detail (example values = Obsolete: Version; Trivial: File Type)

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
Valora's software has considerable capacity to tweak the rules at any time throughout the process. They expect the rules to be iterative in nature as new facets come to light or as client needs change over time. A good example of this are new regulations that require additional or new metadata tagging, such as GDPR.

**What connectivity does the solution have to different data sources, repositories and storage mechanisms?**

*PowerHouse* connects to other data sources via APIs and connectors. Common supported repositories include SharePoint, OpenText, Documentum, and Enterprise Vault, among others. Valora also supports access to other systems that contain data but are not typically considered such as SalesForce, SAP, and UltiPro, as well as cloud-based file storage and sharing systems, such as Box, Dropbox, OneDrive and Google Drive.

*BlackCat* is the central command center (front end to PowerHouse). It is used to manage data and document control as well as reporting and analytics across multiple silos including as shared drives, repository servers, email servers, cloud-based storage, and other data sources.

**Assessment:** Valora Technologies provides AutoClassification, data mining, content analytics, and document intake and visualization. It enables metadata tagging, content extraction and analysis, and rules-

based content disposition across myriad forms of documents, electronic files, data and content. The records management features that include classifying documents for appropriate data retention duration as well as classifying documents for appropriate storage or migration location is useful to our domain.

**Recommendation:** The AA01 study group *recommends* learning more about Valora Technologies' products and services through an interview with a representation who can verify our findings and provide additional information on the AI-features utilized in their products/services.

# Inventory of AI-enabled Software/Services
## Ranking #2
## Not Recommended for Interviews at this Time
## Included in Report Due to Potential Value to Other Researchers and Potential Future Investigation

# AI Enabled Services and Software

**Name of Vendor:**               ANZYZ
**Name of Product / Service:**    Corpus Cube Linguistics (CCL™)
**Self-Described Service:**       Artificial Intelligence Company
**Website Address:**              https://www.anzyz.com/about-us/
**Gartner CSP Magic Quadrant:**   Yes / No
**G2 Grid for ECM:**              Yes / No

**Introduction:**  Anzyz is a Norwegian based Artificial Intelligence company, which integrates supervised, unsupervised and manual rule-based learning through their Natural Language Processing technology, ANZYZ Corpus Cube Linguistics (CCL™).[1]  This AI-drive, text analytics solution for enterprises and public authorities is used within a wide range of industries including Health Care, Energy, Law/Legal and Police/Security.

| AI Features |
| --- |

**AI techniques/approaches:**
- Natural Language Processing (NLP) technology
- text analytics
- search and classify unstructured data
- interpret and understand the contextual meaning in any written natural language without the need to train on large, tagged text collections.

**What functionality does AI enable:**
- Classification of unstructured data
- Natural Language Understanding (contextual meaning)

**Software Offerings (AI-enabled features):**
- *Legal by Anzyz*: Concept builder allows users to target terms in documents, files, and emails. Creates a timeline to visualize trends and occurrences and explore when a concept or combination of concepts appear in the data.
- *Health by Anzyz:* Tool maps patient information in near real time against clinical categories to provide fast and precise quality of care.
- *Well by Anzyz:* Tool to find key data and classify according to topics of interest. Can gain insights into historical drilling and well data.
- *Research by Anzyz:* AI driven text-analytics software to extract insights and discover information.

**How are functions enabled?**
- Train on raw, untagged data.
- XAI: The AI is explainable and contains information about decision-making.
- Products for Legal, health, well, and research
- Legal:
- Explore data, rank for relevance, highlight key elements, label documents from dataset.
- Concept builder
- Visualize trends and occurrences

**What descriptions of algorithms are provided?**
Corpus Cube Linguistics is a text analytical algorithm that is self-learning and enables interpretation and

---

[1] https://www.anzyz.com/about-us/

# AI Enabled Services and Software

understanding of the contextual meaning in Big Data. The algorithm helps users read, analyze, understand, and extract valuable insights from large quantities of complex data. The self-learning algorithm is language neutral, and capable of understanding terminology, jargon, typos, slang and irony.[2]

**Third party services supporting AI?** Anzyz's AI drive Natural Language Processing (NLP) technology is proprietary.

**Opportunity for User input? If rule-based, can it be customized?**
The user's custom-made concepts can be applied on the AI base. This is accomplished from scratch using multiple target-terms in combination with related words put forward by a state-of-the-art AI engine. It is then possible to target all documents, files and emails related to a specific concept.

**What information about the AI models in use is provided/available to the user?**
Customers are informed that Anzyz's solution allows domain experts to build NLP models themselves. The model can be trained on any set of raw (untagged) data. The AI is explainable and contains information about decision making.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
The algorithm used (CCL) integrates supervised, unsupervised, and manual rule-based learning. Workflows can be automated, data analyzed, and key information extracted with off-the-shelf applications.

**Assessment:** This product explicitly mentions AI and bills itself as an AI company. The software's solution analyzes and understands Big Data "from the inside-out" and makes it possible to integrate unstructured and structured data, across information sources. Anzyz CCL™ understands interdependencies and connections in words, sentences, and context – fully based on self-learning from raw data set. The solution is thereby able to understand expressions and characteristics, as well as terminology, misspellings, and abbreviations, in any natural written language. The products are geared to the legal, health care, well and drilling, and research domains. While it would be useful to explore for eDiscovery applications, it is not specifically designed for records management or archives and would not be a priority for further investigation at this time.

**Recommendations:** The AA01 study group *does not recommend* learning more about Anzyz products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services at this time. However, this product is included in the report due to its potential interest to other researchers.

---

[2] https://www.appengine.ai/company/anzyz-technologies

# AI Enabled Services and Software

**Name of Vendor:** AODocs (Altirnao Inc.)
**Name of Product/Service:** AODocs
**Self-Described Service:** Enterprise Content Management Solution
**Website Address:** https://www.aodocs.com/
**Gartner CSP Magic Quadrant:** <span style="color:red">Yes</span>/No
**G2 Grid for ECM:** <span style="color:red">Yes</span>/No

**Introduction:** AODocs is a Niche Player in Gartner's 2021 Magic Quadrant for Content Services and a Leader on the 2022 G2 Grid for Enterprise Content Management. Its CSP, also called AODocs, is built on top of Google Drive and is a multitenant SaaS platform. It provides a full set of CSP capabilities.

## AI Features

**AI techniques/approaches:**
AI uses machine learning and Google cloud to allow the client to create scalable, end-to-end, cloud-based document processing applications.

**What functionality does AI enable:**
AODocs Analytics: Extracts metadata and tags from documents, images, and videos.

**Software Offerings (AI-enabled features):**
Document Management for Google Workspace utilizing Google's Document AI platform.

**How are functions enabled?**
Google's Document AI platform is an Artificial Intelligence-based service that lets the client process documents such as invoices, receipts and forms, regardless of the quality of the source. The connector to the Google Document AI platform extracts specific pieces of data (for example, net amount, VAT and due date from an invoice) and adds them as property values in the AODocs document. The feature works with documents using either Google Drive or Google Cloud storage.

**What descriptions of algorithms are provided?**
The Google Document AI solutions suite includes pre-trained models for data extraction, Document AI Workbench to create new custom models or uptrain existing ones, and Document AI Warehouse to search and store documents.

**Third party services supporting AI?**
Yes, Google Document AI.[1]

**Opportunity for User input? If rule-based, can it be customized?**
The Google Document AI Workbench allows the client to build custom models or uptrain an existing model to meet their business needs. Customization depends on the license. Document AI lets clients train new processor versions using their own training data and evaluate the quality of the processor version against their own test data.

**What information about the AI models in use is provided/available to the user?**

---

[1] https://cloud.google.com/document-ai/docs#docs

# AI Enabled Services and Software

| |
|---|
| No details provided. |
| **Alternatively, if a rules-based system, how much can the user 'tweak' the rules?** <br> N/A |

**Assessment:**  According to Gartner, "AODocs has established itself as a primary consideration for organizations focused on Google Workspace as their new work hub. It is deeply embedded into the Google ecosystem and is therefore able to make extensive use of adjacent capabilities such as Google collaboration and AI."  Google Document AI Workbench allows users to create their own custom models. This product is limited by its dependence on Google Drive.

**Recommendation:**  The AA01 study group *does not recommend* learning more about AODocs products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services at this time. However, this product is included in the report due to its potential interest to other researchers.

# AI Enabled Services and Software

**Name of Vendor:**             Box.com
**Name of Product/Service:**    Content Cloud
**Self-Described Service:**       Secure content management and collaboration
**Website Address:**             https://www.box.com/home
**Gartner CSP Magic Quadrant:**  <span style="color:red">Yes</span>/No
**G2 Grid for ECM:**            <span style="color:red">Yes</span>/No

**Introduction:**   Box employs a lifecycle management approach that includes flexible retention schedules, preservation for defensible discovery, and disposition management.  Two AI offerings are offered: Box Skills applies AI technologies from leading providers and Box Shield identifies regulated data and proprietary information when content is considered "active"—when it's uploaded, shared, previewed, edited, downloaded and more.  It has "security partners" to classify and protect data at scale.

Key features related to retention and disposition are retention policies, event-based retention, legal holds, modifiable retention, advanced trash controls, and unlimited file versions.

| AI Features |
|---|
| **AI techniques/approaches:**<br>*Box Skills:* Use algorithms from providers or bring in your own. Machine learning to manage unstructured content.<br>*Box Shield:* Deep learning and machine learning are used (no reference to specific algorithms). |
| **What functionality does AI enable:**<br>*Box Skills:* Applies best of breed AI technologies from other providers to customer content in box.  Technologies include image labeling, speech-to-text transcription, natural language understanding, and more.<br><br>*Box Shield:* Offers deep learning-based malware detection. Users can configure PII data classification and monitoring or enforcement mode to protect data.  Machine learning produces alerts on potentially comprised accounts and data exfiltration.  See 2020 video demo.  Integration of Box with Box Relay is possible to add classification to content related to business process workflows and Microsoft Office for the web so that Shield's download and security controls can follow the file as being viewed and edited in MS Office online. |
| **Software Offerings (AI-enabled features):**<br>See previous information. |
| **How are functions enabled?**<br>Through algorithms provided by Box (Box Shield) or other vendors (Box Skills). |
| **What descriptions of algorithms are provided?**<br>*Box Skills:* Use algorithms from providers or bring in your own. Machine learning to manage unstructured content.<br>*Box Shield:* Deep learning and machine learning are used (no reference to specific algorithms). |
| **Third party services supporting AI?**<br>*Box Skills:*  Provided by others. |

# AI Enabled Services and Software

| |
|---|
| *Box Shield:* Provided by Box. |
| **Opportunity for User input? If rule-based, can it be customized?**<br>*Box Skills:* Client can use own algorithm<br>*Box Shield:* Client can create and manage custom classification labels, classify content manually or automatically, enforce classification-based security policies. |
| **What information about the AI models in use is provided/available to the user?**<br>Videos are available on website including demos of products. |
| **Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**<br>*Box Skills:* Create own custom integrations to apply AI to your business needs.<br>*Box Shield:* In demo example of potential data leakage and rejection of sharing request with external user using private email address (not corporate). See 2020 video demo. |

**Assessment:**  Box is a single, secure, easy-to-use platform for the entire content lifecycle from file creation and sharing to co-editing, signature, classification, and retention. It is regarded as a leader in the content services platform space by Gartner (2022). Its strengths lie in the areas of ease of use, its strong software development kit (SDK) and API for building content services applications, and prebuilt application integration for other cloud productivity suites including Microsoft (office 365), Google (GSuite) and Apple (iWork).  But the fact that it lags behind other lifecycle management platforms in the areas of records management and workflow make it of limited interest to this study.

**Recommendation:** The AA01 study group *does not recommend* learning more about Box.com products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services at this time. However, this product is included in the report due to its potential interest to other researchers.

# AI Enabled Services and Software

**Name of Vendor:**                 Casepoint
**Name of Product/Service:**      Casepoint legal discovery platform
**Self-Described Service:**          Legal Discovery Platform
**Website Address:**                https://www.casepoint.com/
**Gartner CSP Magic Quadrant:** Yes/No
**G2 Grid for ECM:**               Yes/No
**G2 Grid for eDiscovery:**       Yes/No

**Introduction:**   Casepoint offers a cloud-based, powerful end-to-end platform designed for organizations to manage increasingly complex litigation, investigations, and compliance needs. It aids with eDiscovery, legal holds, investigations, data privacy and compliance, and freedom of information requests.

## AI Features

**What AI techniques/approaches does it use?**

- Advanced analytics
- Active Learning
- Dynamic review workflow
- Advanced search
- Build "data stories" from result of search for key terms, dates.
- Identify near similar documents
- Email threading
- Clustering

**What functionality does AI enable:**

- Identification of relevance, non-relevance, and privilege
- Active learning continuously learns, predicts, and ranks unreviewed documents
- Advanced search using content (text) search, field search, and analytics
- Use clustering to identify and group similar documents.
- Near similar search allows identification of similar documents or culling of documents with duplicate text.

**Software Offerings (AI-enabled features)**:

The product provides an end-to-end solution in 3 steps: Collection and Processing (600+ data types recognized); Review and Production AI-powered active learning (or TAR 2.0) to predict and rank unreviewed documents; Trial Prep and Case Strategy by revealing connections between key facts, documents, data, and people--all the while reducing risk with secure, centralized data and role-based security.

Information on the algorithm was not located but a graphic of the review process was presented:[1]

---

[1] https://www.casepoint.com/ediscovery-platform/artificial-intelligence-advanced-analytics/
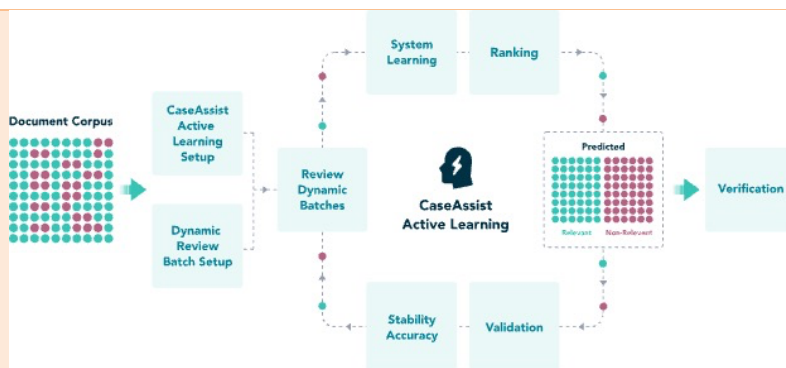
# AI Enabled Services and Software


**Figure 1:** CaseAssist Active Learning Process

**How are functions enabled**?
The Casepoint legal discovery platform provides an end-to-end solution drive by AI technology.

**What descriptions of algorithms are provided**?
A guide on predictive coding[2] (machine learning) can be found on their site.  Casepoint also describes it as Computer-Assisted Review (CAR) or Technology-Assisted Review (TAR). While there are different predictive algorithms and models, the basics are described as:
1. Training Seed Sets: These are created by a group of subject matter experts (SMEs) who tag them as relevant or non-relevant.
2. Prediction Process: Once the software has analyzed the seed set, it creates an internal algorithm capable of predicting future the relevancy of future documents. As more sample documents are added the algorithm is continually refined until the desired results are achieved.
3. Review: The predictive model is then applied to the entire document set so the remaining documents ca be predicted as relevant or non-relevant. Two concepts (not the behind-the-scenes technology) were explained:
   a. Active Learning: an iterative method where the training set is augmented by documents chosen the algorithm and coded manually by the human. It learns from humans to rank unreviewed documents.
   b. Passive Learning: an iterative method using multiple random sample sets for training the algorithm until it achieves the desired results.

Casepoint's Case Assist AI is capable of prioritizing documents for relevance review using dynamic batching. Active learning is used to work in the background and highlight what could be key documents, dates, and people.

**Third party services supporting AI**?
No, this is a proprietary platform. It does, though, provide connectors to review documents in third-party products:  Office 365, Outlook, One Drive, Google Vault, Dropbox, Box, Google Drive, Gmail, Slack, Slack Enterprise, and AutoCAD.

**Opportunity for User input? If rule-based, can it be customized**?
The client can be involved in preparing seed sets and in training the model.

**What information about the AI models in use is provided/available to the user**?

---

[2] https://www.casepoint.com/resources/spotlight/complete-guide-for-predictive-coding-and-how-it-can-be-used-in-ediscovery/#[object%20Object]

# AI Enabled Services and Software

Some information provided as discussed previously.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules**?
N/A

**Assessment:**  While this is not an archives or records/information management product, it will be of interest to information professionals faced with the challenge of responding to requests for electronically stored information (ESI). Casepoint's cloud-based, powerful end-to-end platform can be of use to those organizations faced with eDiscovery, legal holds, investigations, data privacy and compliance, and freedom of information requests.

**Recommendation:** The AA01 study group *does not recommend* learning more about Casepoint products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services at this time. However, this product is included in the report due to its potential interest to other researchers.

# AI Enabled Services and Software

**Name of Vendor:**                   D.velop
**Name of Product/Service:**      D.velop software products
**Self-Described Service:**        Document Management and Digital Business Processes
**Website Address:**               https://www.d-velop.com/
**Gartner CSP Magic Quadrant:** <span style="color:red">Yes</span>/No
**G2 Grid for ECM:**                  Yes/<span style="color:red">No</span>

**Introduction:** D.velop is a visionary in the Gartner Magic Quadrant. It's software primarily focuses on core functionality such as collaborative authoring, external file sharing and electronic signatures. D.velop lacks formal records management capabilities compared to leaders in this market and would require integration with other solutions for features such as file plan management. Product offerings include Enterprise Content Management, Document Management, Invoice Processing, Contract Management and more.

## AI Features

**AI techniques/approaches:**
AI-assisted processing of documents using machine learning and natural language processing.

**What functionality does AI enable:**
AI-based document separation of scanned documents.
Classification of documents based on file type (e.g., invoices, reminders, delivery bills, etc.).
Automatic indexing of contents of documents.

**Software Offerings (AI-enabled features):**
1. *Separate documents*: separate scanned documents (Including multi-page documents)
2. *Classify documents:* artificial intelligence (machine learning) accurately recognizes directly what kind of document it is (e.g., invoice, delivery note). Information can be used to classify documents according to a file plan or route via email to a recipient.
3. *Extract data from documents*: The AI service automatically recognizes patterns in text data and reads out all the information relevant to you using innovative Natural Language Processing (NLP) techniques.
4. *AI enhanced Invoice Processing:* Combination of machine learning and rule-based procedures to improve document and business processes.

**How are functions enabled?**
- Document Separation: Using adaptive text and image analysis, scanned documents (even multi-page documents) are identified.
- Document Classification: AI recognizes the kind of document (ex. invoice), and this information is used later to automatically classify the document in a digital file plan or forward to specific recipients via email.
- Data Extraction from Documents: AI service automatically recognizes patterns in text data and reads all the relevant information. Natural Language Processing (NLP) techniques make it possible (for example) to read content from an invoice without rules set and/or key words.

**What descriptions of algorithms are provided?**
No information could be located.

# AI Enabled Services and Software

| |
|---|
| **Third party services supporting AI?** |
| The solutions appear to be out of the box provided by the vendor. |
| **Opportunity for User input? If rule-based, can it be customized?** |
| The AI was trained on documents already in their repository submitted by users of their products. They describe the user-AI engagement in this way: |
| "We provide you with an 'AI colleague' who reviews document inventories for you in advance, makes recommendations on how to file the documents and helps you to transfer essential information from the documents to the system more quickly. This often works fully automatically. But you retain control at all times and can overrule the AI." |
| **What information about the AI models in use is provided/available to the user?** |
| No information located. |
| **Alternatively, if a rules-based system, how much can the user 'tweak' the rules?** |
| D.velop uses a combination of learning and rule-based procedures. One example is AI Enhanced Invoice Processing described as: "If information such as client, supplier, or invoice data on any invoice is not correctly recognized, it is first improved by the user. The AI then independently creates new recognition logic." |

**Assessment:** D.velop primarily focuses on core functionality such as collaborative authoring, external file sharing and electronic signatures. Its product offerings include Enterprise Content Management, Document Management, Invoice Processing, Contract Management and more. However, D.velop lacks formal records management capabilities and would require integration with other solutions for features such as file plan management.

**Recommendation:** The AA01 study group *does not recommend* learning more about D.velop products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services at this time. However, this product is included in the report due to its potential interest to other researchers.

# AI Enabled Services and Software

| | |
|---|---|
| **Name of Vendor:** | DocuWare |
| **Name of Product / Service:** | DocuWare and DocuWare Cloud |
| **Self-Described Service:** | Document Management Software and Workflow Automation |
| **Website Address:** | https://start.docuware.com/ |
| **Gartner CSP Magic Quadrant:** | Yes / No |
| **G2 Grid for ECM:** | Yes / No |

**Introduction:** DocuWare provides digital management and automated workflows to major industries from manufacturing and retail to healthcare and government. It has been reading content from structured documents for years, but in 2022 it announced the capability of indexing information on unstructured documents and converting it into highly structured, usable data by employing machine learning technology. Docuware is considered a niche player in Gartner Magic Quadrant for Content Services Platforms and a leader for Enterprise Content Management on the 2021 G2 Grid.

| **AI Features** |
|---|
| **AI Techniques/approaches:**<br>Intelligent Indexing can be a part of both DocuWare (on premise as an add-in) and DocuWare Cloud (embedded feature). DocuWare states intelligent indexing is an "AI-based technology converts document content right into indexing terms."[1] Machine learning technology remembers each document and indexing corrections, so every capture increases the speed, accuracy, and reliability of the tool. |
| **What functionality does AI enable:**<br>Intelligent Indexing: The AI used within intelligent indexing can identify valuable information and learn through machine learning which index terms are important as corrections and additions are made by the users. Levels of certainty in indexing is reflected by a color-coding system.[2] |
| **Software Offerings (AI-enabled features):**<br>Intelligent indexing of key terms. |
| **How are functions enabled?**<br>Intelligent Indexing highlights suggested index terms in the document. Depending on the system's level of certainty, they are shown in different colors. Machine learning technology remembers each document and your indexing corrections, increasing the speed, accuracy ,and reliability of the tool. After a few documents, capture is automated and seamless. |
| **What descriptions of algorithms are provided?**<br>No description of the algorithms used could be located. |
| **Third party services supporting AI?**<br>Intelligent indexing appears to be proprietary to the vendor, Docuware. |
| **Opportunity for User input? If rule-based, can it be customized?**<br>Intelligent indexing allows for user input to train its machine learning indexing. |

---

[1] https://start.docuware.com/features-and-capabilities
[2] https://start.docuware.com/docuware-intelligent-indexing?hsLang=en

# AI Enabled Services and Software

**What information about the AI models in use is provided/available to the user?**
No information about the AI models or functionalities are provided for the users.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
N/A

**Assessment:** Docuware has mature capabilities for workflow design and records management planning, but the only area that to date benefits from AI/ML is its intelligent indexing capability.

**Recommendation:** The AA01 study group *does not recommend* learning more about Docuware's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services at this time. However, this product is included in the report due to its potential interest to other researchers.

**Resources:**
DocuWare Cloud white paper (PDF) included in file folder. Can additionally be found here: https://www.docuware.com/main.asp?sig=dld&lan=en&loc=in&dwdblan=english&dwdbkat=do*&dwdbname=white%20paper%20docuware%20cloud.

- Further information about DocuWare and current AI (intelligent indexing feature) and plans for future AI: https://start.docuware.com/blog/product-news/intelligent-indexing-docuware-goes-ai. Published April 27th, 2022.
- Diagram about DocuWare being an ECM: https://start.docuware.com/enterprise-content-management

# AI Enabled Services and Software

**Name of Vendor:**            Grupo Adapting
**Name of Product/Service:**   Abox ECM Software and Services
**Self-Described Service:**    Professional document management solution
**Website Address:**           https://www.adapting.com/en
**Gartner CSP Magic Quadrant:** Yes/No
**G2 Grid for ECM:**           Yes/No

---

**Introduction:** Groupo provides both a document management solution and services that allow the client to co-create AI models using their products. ABOX is the professional document management solution that captures, stores, processes and distributes unstructured business information (documents). Hyperautomation allows the co-creation of machine learning models based on AI.

| AI Features |
|---|
| **AI techniques/approaches:**<br>Groupo Adapting's Hyperautomation is responsible for automating tasks normally performed by human hands. It applies technologies such as Process Modeling (BPA), Robotization (RPA) and Artificial Intelligence (AI) to significantly improve document flows.<br><br>**Hyperautomation involves co-creation of machine learning models based on AI techniques that involve 6 steps: Data collection and sampling.**<br>• Exploratory data analysis.<br>• Variables engineering.<br>• Specification, training, and evaluation of the models.<br>• Calibration of hyper-parameters.<br>• Deployment of the model.[1] |
| **What functionality does AI enable:**<br>• Capture<br>• Automation of Actions with BPM<br>• Detection of Errors and Inconsistencies<br>• Autoclassification with Machine Learning<br>• Sentiment analysis (emotional measurement)<br>• Legal systems integration service |
| **Software Offerings (AI-enabled features):**<br>It is difficult to understand where Hyperautomation is employed in the first 4 of the following Groupo Solutions:<br>• One-stop registration and correspondence management<br>• Records management<br>• Electronic Record<br>• Intranet – Custom Extranet<br>• Hyperautomation (BPA/RPA) |
| **How are functions enabled?**<br>Through Grupo Adapting's Hyperautomation. |

---

[1] https://www.adapting.com/en/hyperautomation/

# AI Enabled Services and Software

| |
|---|
| **What descriptions of algorithms are provided**?<br>Algorithms are proprietary and unavailable for view. |
| **Third party services supporting AI**?<br>No, they appear to be proprietary to Grupo Adapting. |
| **Opportunity for User input? If rule-based, can it be customized**?<br>None or limited other than developed solutions for individual clients. |
| **What information about the AI models in use is provided/available to the user**?<br>Unavailable. |
| **Alternatively, if a rules-based system, how much can the user 'tweak' the rules**?<br>Open-source editable software. Tweakable in  select areas and functions within the proprietary software model.  For example, the functionalities in Hot Folder/HotMailbox allow rules by subdirectories or mailboxes and by file or email attributes. Metadata can be updated, and automation can handle other actions such as page count, file creation, substitution, or creation of secondary attachments. Audit of actions can also be carried out.  It is not clear where AI is employed in this process. |

**Assessment:**  ABOX is a professional document management solution that captures, stores, processes and distributes unstructured business information (documents). It is in use in Spain and Latin America.  Much of the material on the website was in Spanish, including videos. However, there was enough English text to

**Recommendation:**  The AA01 study group *does not recommend* learning more about Groupo Adapting's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services at this time. However, this product is included in the report due to its potential interest to other researchers.

# AI Enabled Services and Software

**Name of Vendor:**　　　　　　　IPRO
**Name of Product/Service:**　　IPRO eDiscovery Suite
**Self-Described Service:**　　　Governance and eDiscovery
**Website Address:**　　　　　　https://ipro.com/
**Gartner CSP Magic Quadrant:**　Yes/No
**G2 Grid for ECM:**　　　　　　Yes/No

**Introduction:**　 IPRO combines governance and eDiscovery software with predictive analytics to give corporations, government agencies, law firms and legal partners an open and effective solution to data challenges. IPRO is positioned in the Leaders category in this *2022 IDC MarketScape* for worldwide eDiscovery early case assessment software, which ranks software on two axes: current capabilities and strategies for the future.

| AI Features |
| --- |
| **AI techniques/approaches:** <br> AI-driven solutions are used to remove human error when reviewing, redacting, and producing documents. |
| **What functionality does AI enable:** <br> • AI assisted document review <br> • Data culling, AI insights, and visual search to reduce data prior to prior to review <br> • Automatically classify data by crawling all unstructured data, regardless of where it's stored. <br> • Automatically identify sensitive personal and health information such as name, contact information, passports, driver license, gender, and much more. <br> • Automatically apply annotations ready for final redaction. <br> • Review data in-place before collection & processing. <br> • Identify and group documents with clustering. |
| **Software Offerings (AI-enabled features):** <br> *Open Discovery*--Connect and review data live with AI Document Review. <br> *Live EDA*—Analyze, search, and review data in place before collection. <br> *ZyLab One*—eDiscovery platform that facilitates legal hold, in place preservation, OCR text extraction, identification of sensitive data, and in-place analytics. |
| **How are functions enabled?** <br> An AI-based classification engine automatically classifies sensitive data, such as PCI, PHI and PII. |
| **What descriptions of algorithms are provided?** <br> None. |
| **Third party services supporting AI?** <br> No. |
| **Opportunity for User input? If rule-based, can it be customized?** <br> Yes. |
| **What information about the AI models in use is provided/available to the user?** |

# AI Enabled Services and Software

> None.

> **Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
> No information but it does tie to the organization's retention schedule, so input from the organization is required.

**Assessment:**  Review continues to be the costly aspect of eDiscovery despite technological advances. IDC Marketplace's vendor evaluation is based on a comprehensive criterion expected to be most conducive to success in providing tools to enable collecting, processing, culling, and analyzing data in conjunction with investigations and litigations. Although not an enterprise content management system nor a content services platform, this product has capabilities that are important to records and information managers, such as deduplication, identification and removal of ROT, document classification, in-place content analysis, and management of retention requirements and legal holds.

**Recommendation:** The AA01 study group *does not recommend* learning more about IPRO's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services at this time. However, this product is included in the report due to its potential interest to other researchers.

# AI Enabled Services and Software

**Name of Vendor:**      Merlin.tech
**Name of Product/Service:**      Sherlock Integrated Search
**Self-Described Service:**      AI-Power Discovery; Cloud Native Software
**Website Address:**      https://www.merlin.tech/
**Gartner CSP Magic Quadrant:**      Yes/No
**G2 Grid for ECM:**      Yes/No

**Introduction:**      Merlin is a cloud technology company developing AI-powered software to help professional teams manage investigations, discovery, early case assessment and other matters involving large quantities of documents. Merlin's cloud-native architecture is built to exploit the scale, elasticity, and flexibility of the AWS cloud.

| AI Features |
| --- |

**AI techniques/approaches:**
Merlin combines AI, search, and secure cloud technologies with their proprietary software code to make search smarter, security more effective, and discovery less costly.

**What functionality does AI enable:**
Search, analytics, and review of documents

**Software Offerings (AI-enabled features):**
Sherlock Integrated Search facilitates:
- Keyword Search
- Freeform search
- Clustering

**How are functions enabled?**
Algorithmic keyword search refined with machine learning.

**What descriptions of algorithms are provided?**
*Sherlock Integrated Search*—Three ways to work with the search engine: 1) Simple keyword search (using quotes around phrase) with approval of those that are returned that are relevant and dismissal of those tagged not relevant. As engine learns what we approve and don't, it builds a new model to become more refined.  2) *Freeform search capability* with brackets around text. Copy positive documents to a folder for review later. You can send documents to Sherlock  When 'submit' is selected, Sherlock creates an AI model, analyze all documents on the site, rank according to relevance, and then return most likely relevant with keyterms. Can create Word Cloud and list view of top 30 terms.  Can boost or suppress terms. Thumbs up bottom is used to tag as a relevant document.  3) Cluster batching—cluster documents by similarity of content. Can tag documents or entire cluster. When submit those reviewed and Sherlock will make a new model.

**Third party services supporting AI?**
None.

**Opportunity for User input? If rule-based, can it be customized?**
Yes. Can help train model using thumbs up and thumbs down buttons to indicate relevance, non-relevance.

# AI Enabled Services and Software

| | |
|---|---|
| **What information about the AI models in use is provided/available to the user?**<br>Quick Guides and Videos. | |
| **Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**<br>N/A | |

**Assessment:**  Merlin provides software to fit organizations' compliance and legal process needs. It leverages algorithms to create models based on documents presented to it and tags created by client indicating relevance or non-relevance.  According to a statement on the website, "Sherlock is the first search, analytics, and review platform to combine the familiarity of keywords with the power and speed of AI. It's a faster, easier, and more effective way to find relevant documents for investigations, discovery, research, and regulatory compliance." [1]  This product would be interesting to investigate further if our focus were eDiscovery.

**Recommendation:** The AA01 study group *does not recommend* learning more about Merlin's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services at this time. However, this product is included in the report due to its potential interest to other researchers.

---

[1] https://www.merlin.tech/software/revieweffective/

# AI Enabled Services and Software

**Name of Vendor:**           Micro Focus International
**Name of Product/Service:**  Content Manager, Control Point
**Self-Described Service:**   Offer Mission Critical Software and Services
**Website Address:**          https://www.microfocus.com/en-us/home
**Gartner CSP Magic Quadrant:** Yes / No
**G2 Grid for ECM:**          Yes / No

**Introduction:**  Micro Focus is one of the world's largest enterprise software providers delivering mission-critical technology and supporting services to customers worldwide.

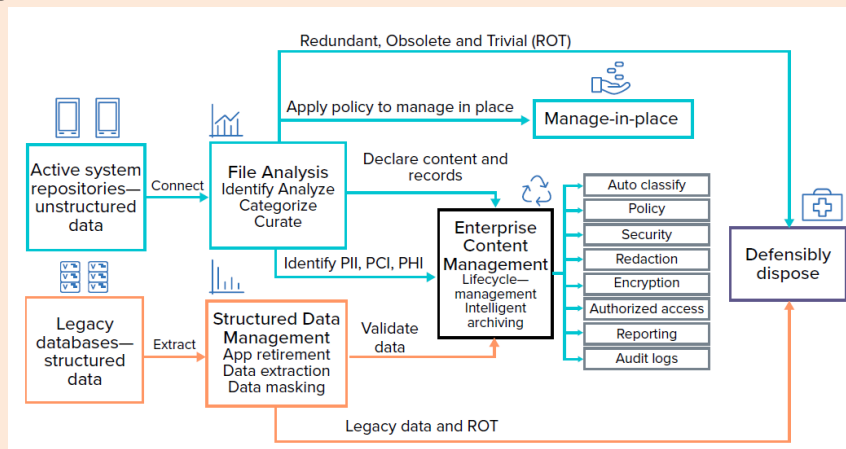| AI Features |
| --- |
| **AI techniques/approaches**:<br>Micro Focus Software-as-a-Service (SaaS) AI technologies are employed to analyze, categorize, and auto-classify unstructured data; extract and mask structured data; and facilitate lifecycle management and intelligent archiving. |
| **What functionality does AI enable**:<br>Micro Focus offers a unified, high-performance advanced analytics and machine learning platform in an AWS cloud account. |
| **Software Offerings (AI-enabled features)**:<br>The Secure Content Management suite illustrated in Figure 1 includes three products:<br>    1.  Control Point – file analysis for unstructured data;<br>    2.  Structured Data Manager – application retirement and structured data archiving); and<br>    3.  Content Manager – governance-based ECM to manage un/structured content across the lifecycle.<br><br>**Figure 1:** Secure Content Management suite |
| **How are functions enabled**?<br>AI functions include analyzing video, image, audio, and text data in real time for "privacy regulation compliance," "sentiment analysis," "deep insights," and "surveillance" of media, web content, social media, etc.  Details of the ML process built into the product are shown in Figure 2. |

# AI Enabled Services and Software
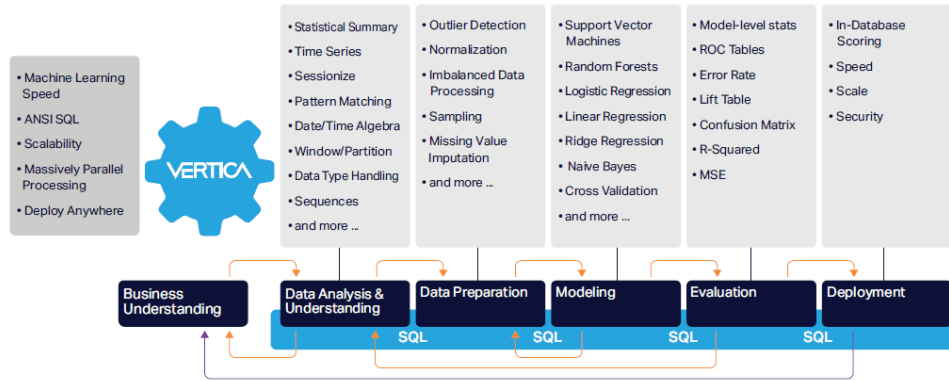


**Figure 2.** Vertica Built-in Machine Learning Process Flow

**What descriptions of algorithms are provided**?
See Figure 1 for information on ML Modeling using algorithms such as Random Forests and Naïve Bayes.

**Third party services supporting AI**?
Micro Focus products are built on intelligent data operating layer (IDOL) software "for advanced search, knowledge discovery, and analytics of unstructured data."

**Opportunity for User input? If rule-based, can it be customized**?

Where a client uses products from the Secure Content Management suite (Figure 1), there is an out-of-the-box aspect. It is difficult to tell how much input the expert users have.

In the case where a client is building using the Vertica data analytics platform, they have a great deal of input, including which models to use and in which order (Figures 2 and 3).
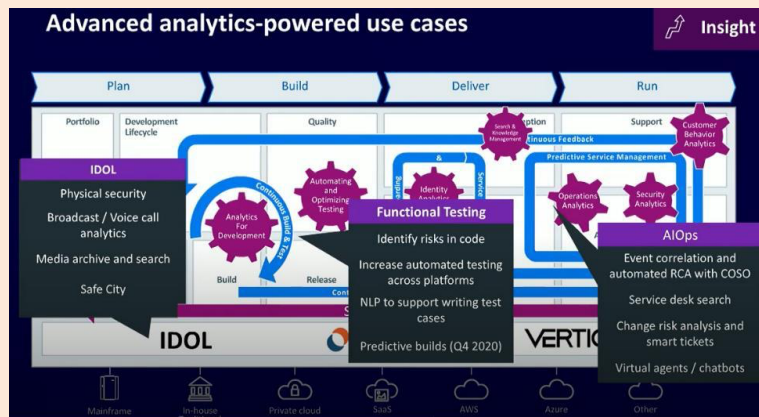


**Figure 3:** Analytics Powered Use Cases

**What information about the AI models in use is provided/available to the user**?
See Figures 1-3 above.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules**?
See above.

# AI Enabled Services and Software

**Assessment:**  Micro Focus Enterprise Software and Services provide the following solutions: Application Delivery, Application Modernization, IT Operations, AI Data Analytics and Governance, CyberRes (resilience), and Vertica (for analytics).  The service most closely related to archives and records management tasks is Micro Focus Filr, listed on the G2 Grid for Enterprise Content Management solutions. However, its strength is the areas of file sharing and file editing like Dropbox and Google Drive.

**Recommendation:**  The AA01 study group *does not recommend* learning more about Micro Focus products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services at this time. However, this product is included in the report due to its potential interest to other researchers.

**Resources**

Figure 1 is from Micro Focus' 2020 brochure "Bringing Information to Life"
https://www.microfocus.com/media/brochure/bringing_information_to_life_and_use_it_brochure.pdf

Figure 2 is from a CyberRes Reduce Risk and Protect Privacy across Enterprise Systems solution flyer
https://www.microfocus.com/media/flyer/reduce_risk_and_protect_privacy_across_enterprise_systems_flyer.pdf

Figure 3 is from the second of three "Smart Digital Transformation" webinars.
1) Run, Transform, and Adapt; 2) AI and ML for Smart DX; 3) Run and Transform Together

# AI Enabled Services and Software

| | |
|---|---|
| **Name of Vendor:** | Smarsh |
| **Name of Product/Service:** | Smarsh Solutions: Enterprise Platform, Compliance Management, Cybersecurity & Risk Management, Digital Safe |
| **Self-Described Service:** | Enterprise Information Archiving (communications capture & archiving) |
| **Website Address:** | http://www.Smarsh.com |
| **Gartner CSP Magic Quadrant:** | Yes/No |
| **G2 Grid for ECM:** | Yes/No |

**Introduction:**   Smarsh is recognized as a leader in the Gartner Magic Quadrant for Enterprise Information Archiving 2022.  Smarsh captures communications from email, social, mobile, IM & collaboration, and voice channels. Its acquisition of Digital Reasoning expands Smarsh's offerings with new specialized solutions utilizing natural language processing (NLP) and machine learning for conduct surveillance and healthcare automation. This acquisition may be a roadmap for future AI features.

## AI Features

**AI techniques/approaches:**

Smarsh offers four cloud-based platforms: Enterprise Platform, Compliance Management, Cybersecurity and Risk Management, and Digital Safe (archive). The Smarsh Enterprise Platform uses elastic compute, natural language processing, machine learning, and artificial intelligence to capture, retain, analyze, and act on the signals in communication data.

**What functionality does AI enable**: The Enterprise Platform delivers capabilities across four areas: improved capture, intelligent data lifecycle management, enhanced machine learning for regulatory compliance, and secure cloud architecture to minimize threats.

**Software Offerings (AI-enabled features)**:
*Enterprise Platform*: Capture, retain, analyze, and act on the signals in your communications data.  Elastic compute (resizable compute capacity in the cloud—user can create, launch, terminate as needed paying by the second for active servers), natural language processing, machine learning, and artificial intelligence.  Capture of email, social, mobile, IM & collaboration, and voice channels.
*Compliance Management*:  Email hosting, web archiving, e-discovery, unify cybersecurity & compliance.
*Cybersecurity & Risk Management*: Cybersecurity & vendor risk
*Digital Safe*: Archiving/discover/compliance/audit tasks managed here.

**How are functions enabled**? This depends upon the solution.  In 2020 Smarsh introduced Echo Cancellation that works specifically for Microsoft Exchange and Office 365 email to handle repetitive messages forwarded along with another communication (response or forwarded message).  Echo Cancellation uses an algorithm that can analyze the conditions of an alert and make intelligent decisions to identify false positives. It then subordinates the review of these false positives within queues, which can be configured by the client. A pilot of this solution demonstrated a 20% reduction in items requiring review with 100% precision (identifying an echo as an echo – true positive) as opposed to identifying no-echoes as an echo (false positive).

**What descriptions of algorithms are provided**?
See above for one example.

# AI Enabled Services and Software

**Third party services supporting AI**?
No, these are Smarsh solutions.

**Opportunity for User input? If rule-based, can it be customized**?
The client can configure policies to analyze its communications using NLP, ML, AI, and refined lexicons.

**What information about the AI models in use is provided/available to the user**?
No information about the models themselves.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules**?
N/A

**Assessment:** This is a solution for digital communications capture, retention, analysis, and use to make decisions and take actions. While not specific to our needs for general enterprise content management, communications must be retained for various reasons including compliance with laws and regulations. Smarsh's products and services will be of interest to most information professionals

**Recommendation:** The AA01 study group *does not recommend* learning more about Smarsh's products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services at this time. However, this product is included in the report due to its potential interest to other researchers.

# AI Enabled Services and Software

**Name of Vendor:**              Xerox
**Name of Product/Service:**    Xerox DocuShare Go
**Self-Described Service:**      Content Management Platform
**Website Address:**           https://www.xerox.com
**Gartner CSP Magic Quadrant:** Yes / No
**G2 Grid for ECM:**              Yes / No

**Introduction:** Xerox has three products to assist organizations manage, automate, and share content. DocuShare is the classic Enterprise Content Management Solution that allows stores all files in one place for collaboration. It is ranked as a Leader on the G2 Grid for ECM. DocuFlex is the cloud-based, SaaS solution explained as the "next generation" platform with all the features of classic DocuShare, which can be managed on site. DocuShare Go Content Management Platform is Xerox's newest offering--a cloud hosted Software as a Solution (SaaS) collaborative Content Services Platform that employs artificial intelligence (AI) and machine learning (ML) technologies. This review is of the AI-enabled offering: Xerox DocuShare Go.

| AI Features |
|---|
| **AI techniques/approaches:**<br>AI recognizes captured content, tags it, and stores it without manual input or intervention. Approval workflows automate notification requests to validate documents, then route them to defined folders based on acceptance or rejection decisions. |
| **What functionality does AI enable:**<br>• Data Extraction using AI and ML.<br>• Content analysis (e.g., account number, logo, name, signature).<br>• Autocategorization (e.g., accounts payable, completed forms, special projects, expense reports)<br>• Automatic tagging (e.g., account number). The feature works with handwritten notes, text within images, and logos.<br>• Automatically triggered workflows (approval needed) |
| **Software Offerings (AI-enabled features):**<br>See above. |
| **How are functions enabled?**<br>ML technology constantly updates its database, ensuring new data is remembered to speed up processing. |
| **What descriptions of algorithms are provided?**<br>None. |
| **Third party services supporting AI?**<br>"DocuShare Go uses Amazon Web Services to deliver the highest standards for privacy, compliance, and data security."[1] |
| **Opportunity for User input? If rule-based, can it be customized?** |

---

[1] https://fccinnovation.co.uk/news/xerox-docushare-go

# AI Enabled Services and Software

Yes, users can add or customize pre-set document types and manage document type fields to allow additional metadata. Users can also set expiry date for disposition.

**What information about the AI models in use is provided/available to the user?**
None.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
Users can add new or customize pre-set document types. Document type fields allow for additional metadata. Users can set expiry dates to manage document lifecycle and compliance initiatives.

**Assessment:** Xerox has a solid reputation for its performance in the Enterprise Content Management field with its DocuShare classic offering as evidenced by inclusion on the G2 Grid for ECM. Xerox DocuShare Go is the latest offering (launched March 29, 2022) designed as an affordable cloud-based software-as-a-service (SaaS) content management platform for small and mid-sized businesses. The information available is sparce, but the fact It states it uses AI and ML to extract data, analyze content, and automatically tag files for content storage, file retrieval, and expiry makes it of interest to our retention and disposition study.

**Recommendation:** The AA01 study group *does not recommend* learning more about Xerox DocuShare Go through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services at this time. However, this product is included in the report due to its potential interest to other researchers.

## Inventory of AI-enabled Software/Services
## Ranking #3
## Not Recommended for Interviews
## Included in Report Due to Potential Value to Other Researchers

# AI Enabled Services and Software

**Name of Vendor:**                       Amazon AWS
**Name of Product / Service:**      Amazon SageMaker
**Self-Described Service:**          Cloud Machine Learning Platform
**Website Address:**                 https://aws.amazon.com/sagemaker/
**Gartner CSP Magic Quadrant:**   Yes / No
**G2 Grid for ECM:**              Yes / No

**Introduction:** Amazon SageMaker is a cloud machine-learning platform launched in November 2017 that enables developers to create, train, and deploy machine-learning (ML) models in the cloud or on embedded systems and edge-devices. In addition to data scientists and developers, business and data analysts are enabled to build, train, and deploy machine learning models for any use case with fully managed infrastructure, tools and workflows.

| AI Features |
| --- |
| **AI techniques/approaches:**<br>This machine-learning platform provides a choice of tools through an integrated development environment (IDE) for data scientists and developers and a no-code interface for business and data analysts to build, training, and deploy machine learning (ML) models. |
| **What functionality does AI enable:** ML; Use ready-made, purpose-built AI services, or your own models with AWS ML services. |
| **Software Offerings (AI-enabled features):** Examples of offerings available:<br>• Amazon SageMaker Canvas provides tools to generate accurate predictions for business analytics without writing code or requiring ML experience.[1]<br>• Amazon SageMaker for Data Scientists is an integrated development environment (IDE) for the ML Lifecycle. SageMaker for Data Scientists offers a broad set of ML capabilities to access, analyze data, and build, train, and deploy high-quality ML Models. Data can be accessed from structured and unstructured data sources.[2]<br>• Amazon SageMaker for ML Engineers provides purpose-development tools for machine learning operations (MLOps).[3] Using the purpose-b uilt tools for ML Operations, processes can be automated and standfardized across the ML lifecycle making it easy to train, test, troubleshoot, deploy, and govern ML models at scale. |
| **How are functions enabled?**<br>● Solve common business problems with AI<br>    ○ Solutions to help enhance customer experiences, enable faster and better decision-making, and optimize business processes.<br>    ○ Explore AI use cases »<br><br>● Add AI to your business applications<br>    ○ Add AI to your business applications |

---

[1] https://aws.amazon.com/sagemaker/business-analyst/
[2] https://aws.amazon.com/sagemaker/data-scientist/
[3] https://aws.amazon.com/sagemaker/mlops/

# AI Enabled Services and Software

- o Purpose-built AI services for speech, vision, documents, and more so developers can get started without any ML experience.
- o View AI services »

- • Build, train, and deploy ML models for any use case
  - o Fully managed infrastructure, tools, and workflows for data scientists and ML developers.

**What descriptions of algorithms are provided?**

Amazon SageMaker Canvas to build ML models without writing code following the process shown in figure 1
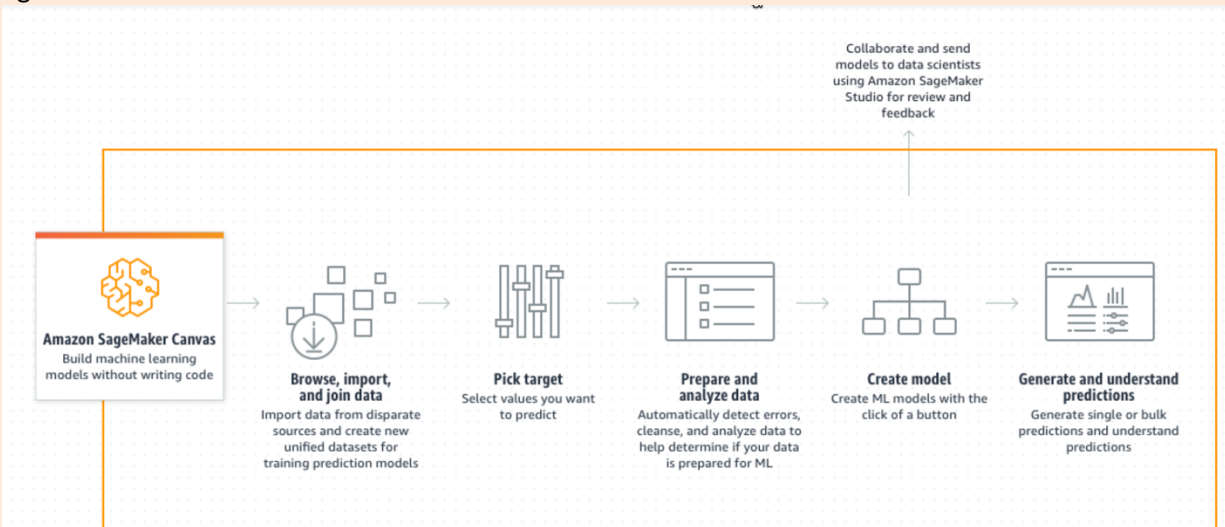


**Figure 1.** Amazon SageMaker Canvas process to generate predictions using ML.
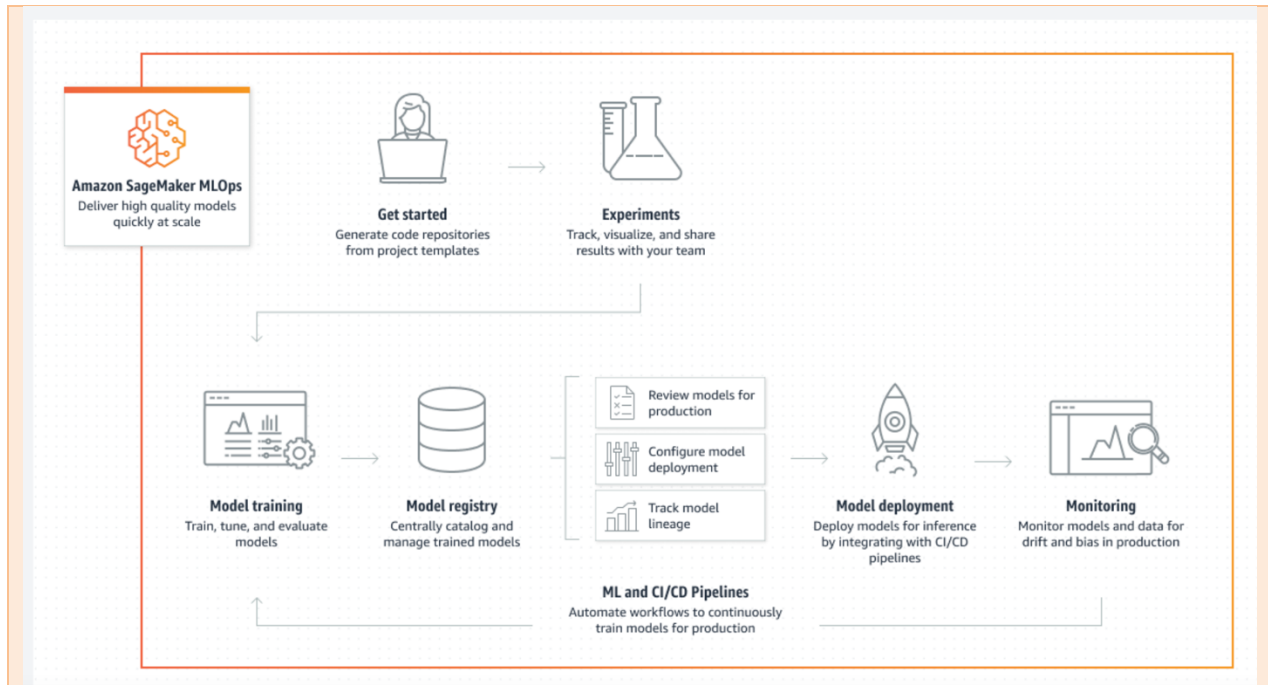
Amazon SageMaker for Data Scientists (IDE) process is shown in figure 2.
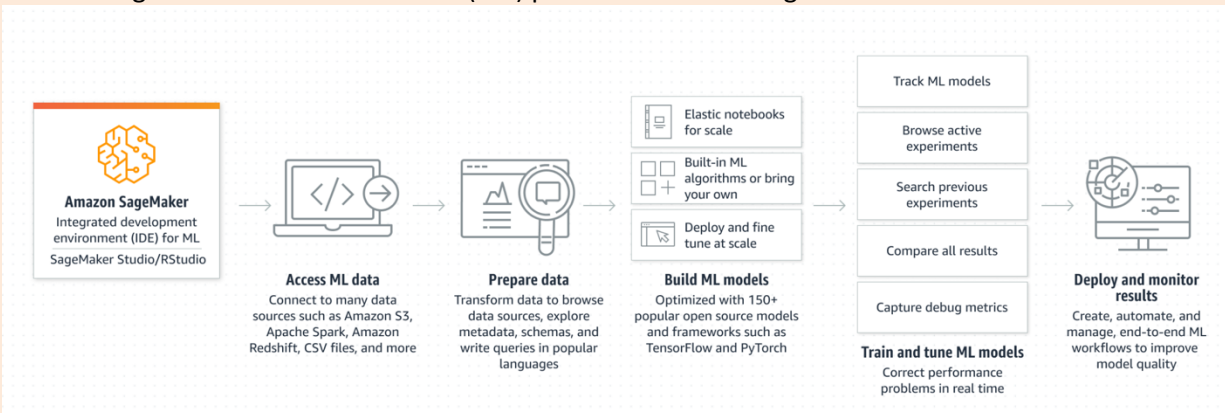


**Figure 2.** Amazon SageMaker IDE process for Data Scientists.

Amazon SageMake MLOps works as shown in figure 3.

# AI Enabled Services and Software



**Figure 3.** Amazon SageMaker MLOps.

**Third party services supporting AI?**
None mentioned.

**Opportunity for User input? If rule-based, can it be customized?**
Users can train and review ML

**What information about the AI models in use is provided/available to the user?**
Information is provided on the website.  A number of ML tutorials are provided based on role (e.g., data scientists, ML engineers, business analysts). A customer enablement web page provides links to additional resources.[4]

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
N/A

**Assessment:** AWS is a noted leader in AI & ML with enough capabilities to offer AI and ML as a service to multiple vendors.  Amazon SageMaker allows the development, training, and deployment of AI ML models. It is not a product or service developed to manage data/content/records.

**Recommendations:** The AA01 study group *does not recommend* learning more about Amazon SageMaker through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services. However, this product is included in the report due to its potential interest to other researchers.

---

[4] https://aws.amazon.com/customer-enablement/?nc2=h_ql_ce_cust-a

# AI Enabled Services and Software

**Name of Product:** Fabasoft AG
**Name of Product Service:** Fabasoft PROCECO
**Self-Described Service:** Curated Ecosystem for document-intensive business processes
**Website Address:** https://www.fabasoft.com/en
**Gartner CSP Magic Quadrant:** <span style="color:red">Yes</span> / No
**G2 Grid for ECM:** Yes / <span style="color:red">No</span>

**Introduction:** Fabasoft is recognized as a niche player in Gartner's 2021 Magic Quadrant for Content Services Platforms. Document management and business process management take place in Fabasoft's Business Process Cloud.

| AI Features |
|---|
| **AI techniques/approaches**:<br>On AI-enabled feature of Fabasoft Proceco is Contract Management where relevant data is extracted from contracts.[1] The user can specify customized contract fields.  The Contract Management software was developed by Fabasoft and Mindbreeze (its subsidiary). The eGov Suite also employ Fabasoft Mindbreeze AI technology for semantic searches. |
| **What functionality does AI enable**:<br>AI-Components<br><ul><li>Digitizing contracts & fill in user fields</li><li>Semantic full text search with filter options</li></ul> |
| **Software Offerings (AI-enabled features)**:<br>*Semantic search* is enabled through Fabasoft Mindbreeze—a link on the Fabasoft Website presents the Mindbreeze Inspire website. Fabasoft Mindbreeze is described as "semantic analysis and machine learning technology that can extract metadata and interpret unstructured data and search queries. Data can be queried intuitively due to the use of National Language Question Answering (NLQA) dialogue.<br><br>*Fabasoft Contracts*, an AI-based contract management software, applies neural networks and machine learning methods to understand key contractual content and clauses. |
| **How are functions enabled**?<br>The AI-based Fabasoft *Contracts software* applies neural networks and machine learning methods to understand key contractual content and clauses. OCR technology makes scanned documents readable and, thanks to an intelligent semantic full-text search, the scanned documents can be located on the basis of their content.<br><br>*Semantic search* using Fabasoft Mindbreeze employs semantic analysis and machine learning technology to extract metadata and interpret unstructured data and search queries. In a Natural Language Question Answering (NLQA) dialogue, data can be queried intuitively. |
| **What descriptions of algorithms are provided**?<br>See previous question and answer. |
| **Third party services supporting AI**? |

---

[1] https://www.fabasoft.com/en/products/fabasoft-contracts

# AI Enabled Services and Software

The AI employed was developed by Fabasoft and its subsidiary, Mindbreeze.

**Opportunity for User input? If rule-based, can it be customized?**
Opportunity for user input was not clear.

**What information about the AI models in use is provided/available to the user?**
Limited information available.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
N/A

**How is this product/service used by organizations to manage content/information?**

**Assessment:**
Fabasoft platform for document intensive business processes employs the Mindbreeze Insight Engine for AI. Due to the use of Mindbreeze AI, the Fabasoft Contract Management software was nominated in 2021 for the eAward in the category of Machine Learning and Artificial Intelligence. Fabasoft is used primarily in the European market. It does focus on content and records management features such as including labeling, retention, and disposition.

**Recommendations:** The AA01 study group *does not recommend* learning more about Fabasoft products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services. However, this product is included in the report due to its potential interest to other researchers.

# AI Enabled Services and Software

**Name of Vendor:**               iManage
**Name of Product/Service:**     iManage
**Self-Described Service:**        Knowledge work platform
**Website Address:**             https://imanage.com/
**Gartner CSP Magic Quadrant:**  Yes/No
**G2 Grid for ECM:**             Yes/No

**Introduction:** iManage is listed on the Gartner CSP Magic Quadrant 2021 as a Niche Player and the G2 Grid for ECM 2021 as a high performer. iManage combines artificial intelligence, security, and risk mitigation with document and email management. iManage focuses its CSP, named iManage Work, on business areas where case work and formal document production are primary use cases. This is most common in legal, but also includes accounting, professional services and financial services.

| AI Features |
| --- |
| **AI techniques/approaches:** <br> AI (machine learning) and contextual search combined to gather insights from large volumes of documents, contracts, and business communications. |
| **What functionality does AI enable:** <br> According to website, "Embedded artificial intelligence automates mundane tasks to streamline processes and enable smarter decisions."[1] |
| **Software Offerings (AI-enabled features):** <br> iManage Extract: <br> • It identifies and extracts information from legal documents, creating structured data. <br> • It combines AI and search to improve visibility into legal risk within contracts. |
| **How are functions enabled?** <br> Extract's 5 key features: <br> • *Collect* by connecting to Work 10 and other document repositories or upload using drag and drop. <br> • *Cluster and Classify* documents based on content using machine learning. <br> • *Extract* key critical data points from documents using the Extractor Library or Extractor Toolbox. <br> • *Workflows* can be created to manage training, review, and quality assurance processes. <br> • *Search and export* by conducting bulk searches and connecting to external systems to export data, crate dashboards, and populate documents. |
| **What descriptions of algorithms are provided?** <br> No information discovered. |
| **Third party services supporting AI?** <br> The solution appears to be proprietary to iManage. |
| **Opportunity for User input? If rule-based, can it be customized?** |

---

[1] https://imanage.com/product-overview/platform/

# AI Enabled Services and Software

The Extractor Toolbox can be used to train new extractors. The tools leverage machine learning and regular expression.

**What information about the AI models in use is provided/available to the user?**
No information was found on their website. However, iManage created the US AI University to educate customers on which tools to use for specific cases and how to create more effective rules-based models.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
Workflows can be used from the Extractor library and new workflows created. The Extractor toolbox can be used to train new extractors.

**Assessment:** iManage offers several products. It is not clear how the AI product, Extract, is used with the one most relevant to records and information management: *iManage Work* for intelligent document and email management. This product allows search across different platforms. In one case study, users in a law firm were pleased with the speed with which content could be located in their most used application, Outlook. The product can be integrated with other productivity suite tools like Microsoft 365, Word, Teams, Gmail, and Google Workspace.

**Recommendation:** The AA01 study group *does not recommend* learning more about iManage products and services through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services. However, this product is included in the report due to its potential interest to other researchers.

# AI Enabled Services and Software

**Name of Vendor:**         Microsoft
**Name of Product:**        Azure AI
**Self-Described Service:**     AI solutions for developers and data scientists
**Website Address:**        https://azure.microsoft.com/en-us
**Gartner CSP Magic Quadrant:**  Yes / No
**G2 Grid for ECM:**         Yes / No

**Introduction:** Microsoft's Azure AI is included on the 2022 Gartner Magic Quadrant for Cloud Infrastructure and Platform Services. Azure AI is comprised of *Azure Applied AI Services*, *Azure Cognitive Services, Azure Machine Learning,* and *AI Infrastructure.* These tools can be used modernize business processes; add AI capabilities to their apps; build, training, and deploy machine learning models; and take advantage of large-scale AI infrastructures and innovative training tools.

| AI Features |
| --- |
| **AI techniques/approaches:** <br> Azure AI is a portfolio of artificial intelligence (AI) services designed for developers and data scientists— to build and deploy their own AI solutions. Access high-quality vision, speech, language, and decision-making AI models through simple API calls, and create machine learning models with tools like Jupyter Notebooks, Visual Studio Code, and open-source frameworks like TensorFlow and PyTorch.[1] |
| **What functionality does AI enable:** <br> Development of AI solutions based on customizable pretrained models. |
| **Software Offerings (AI-enabled features):** <br><br> **Azure Applied AI Services** can be applied to specific business scenarios to: <br> • Modernize business processes with tasks-specific AI. <br> • Accelerate development with built-in business logic that enables solutions to be launched in days, not months. <br> • Run responsibly due to security that extends from the cloud to the intelligent edge. <br><br> **Azure Cognitive Services** allow data scientists and developers with customizable, pretrained models to: <br> • Improve customer experiences with Cognitive Service for Speech (speech to text, text to speech, speech translation, and speech recognition). <br> • Understand conversations with unstructured text with Cognitive Service for Language (Entity recognition, sentiment analysis, question answering, conversational language understanding, and language translation). <br> • Identify and analyze content within images and videos (computer vision, custom vision, Face API). <br> • Make Smarter Decisions (anomaly detector, content moderator, personalizer). <br> • Leverage dynamic language models to power apps (Open AI Service providing advanced coding and language models that can be applied to a variety of use cases). <br><br> **Azure Machine Learning** provides an end-to-end platform for building, training, and deploying machine learning models. Designers and data scientists can: |

---

[1] https://azure.microsoft.com/en-us/solutions/ai/#overview

# AI Enabled Services and Software

- Develop with a choice of tools and automated machine learning.
- Create and deploy models at scale with automated and reproducible machine learning workflows.
- Innovate with a rich set of built-in capabilities to help you understand, protect, and control data, models, and processes.
- Build with best-in-class support for open-source frameworks and languages.

**AI infrastructure** and innovative training tools:
- Allow access large-sale infrastructure.
- Enable hybrid and multi-cloud deployments.
- Access a breadth of AI hardware.

**How are functions enabled?**
In terms of the 'out of the box' models you use when using some of these services, there does seem to be information on the versioning of the models, including Azure Cognitive Services.

**What descriptions of algorithms are provided?**
Azure Machine Learning allows data scientists and developers to build, deploy, and manage models. Support is provided for the end-to-end machine learning lifecycle.  Specific models were not located.

**Third party services supporting AI?**
Azure AI is a Microsoft service.

**Opportunity for User input? If rule-based, can it be customized?**
N/A

**What information about the AI models in use is provided/available to the user?**
Microsoft provides a great deal of information online, including research products, webinars, and tutorials.

**Alternatively, if a rules-based system, how much can the user 'tweak' the rules?**
N/A

**Assessment:** Microsoft's content services platform is considered a leader by Gartner.  Those are the focus of a separate evaluation and a recommendation to explore them further was made.  Microsoft is also highly regarded for its Azure AI Platform, which is a portfolio of AI services designed for developers and data scientists to add speech, vision, language, and decision capabilities to enterprise applications. Azure AI is not a product or service developed to manage data/content/records.

**Recommendations:** The AA01 study group *does not recommend* learning more about Microsoft's Azure AI through an interview with a representative who can verify our findings and provide additional information on the AI-features utilized in their products/services. However, this product is included in the report due to its potential interest to other researchers.

# Inventory of AI-enabled Software/Services
# Summary

This report was compiled based on publicly available information examined by various researchers. The details gathered were not reviewed by the vendors of the products and services investigated. The findings reflect what a typical individual might determine by reviewing materials available in a limited amount of time without input from the entities developing the products/services. Of the initial 56 vendors/products recommended, the number was reduced to 41. The results of a further review of the 41 led to 25 that have been recommended for the interview stage of the study. The forms compiled for the 41 are included in this report; however, only the 25 recommended for the interview stage will be investigated further during the next phase of this study, the interview phase.

Before moving on to that phase, however, it may be useful to highlight some findings from the top 25 vendor products/services identified in phase 1 as incorporating some type of AI functionality in their offerings.

**Branch of Artificial Intelligence**

The team wondered what types of AI were referenced most often in publicly available materials. All 25 of the vendor products/services recommended used the term AI. As shown in Figure 2, according to the researchers describing the offerings, 23 of them explained the use of Machine Learning (ML); 11 referred to Natural Language Processing (NLP); 5 referred to Deep Learning (DL); and 2 referred to Natural Language Understanding (NLU).
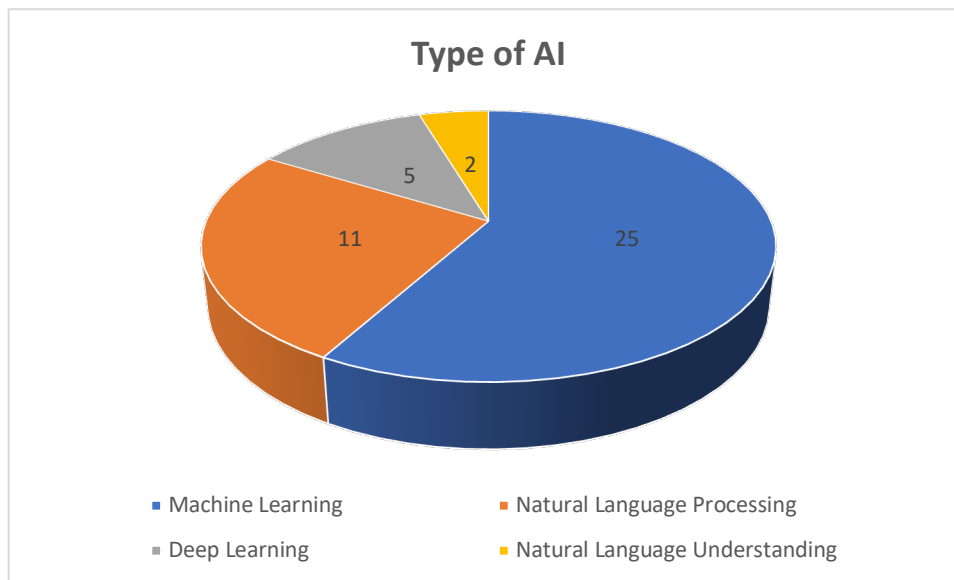


**Figure 2:** Type of AI referenced in materials reviewed for top 25 vendor products/services.

**Term used to describe focus of product/service**

Archivists and records managers are primarily concerned with the management of records. However, increasingly their responsibilities include managing data and information. In addition, developers of products that can be used to manage records use terms other than records, such as content and documents. Some specifically mention the management of email. Recordkeepers are encouraged to include those products and services that may omit the term "records" but that may house records that need to be managed.  As shown in Figure 3, according to the researchers describing these offerings, only 19 of the 25 vendor products/services mentioned records in their materials, while 23 used the term content and 22 the term documents. Of the 25 vendor products/services, 12 stated they could manage email, which may or may not contain records.
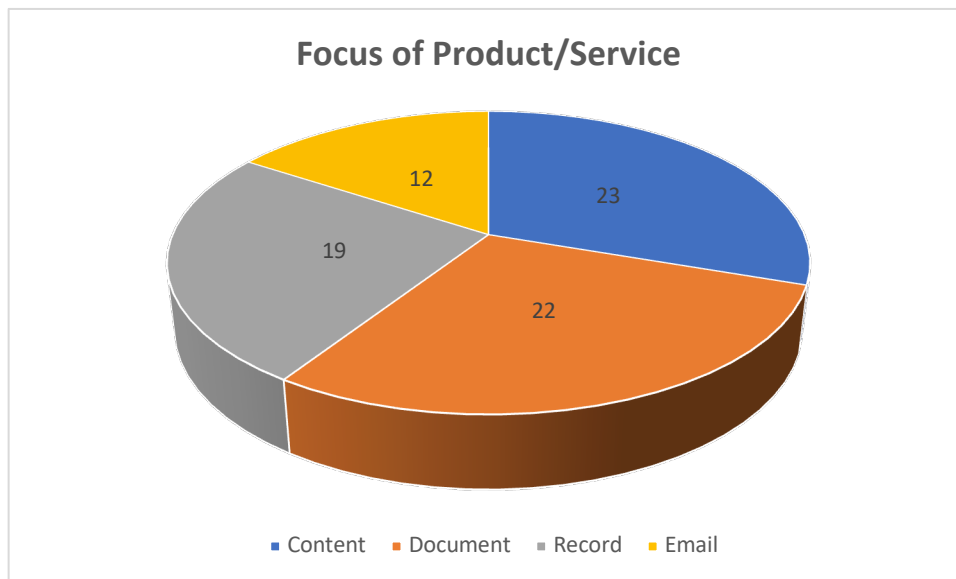


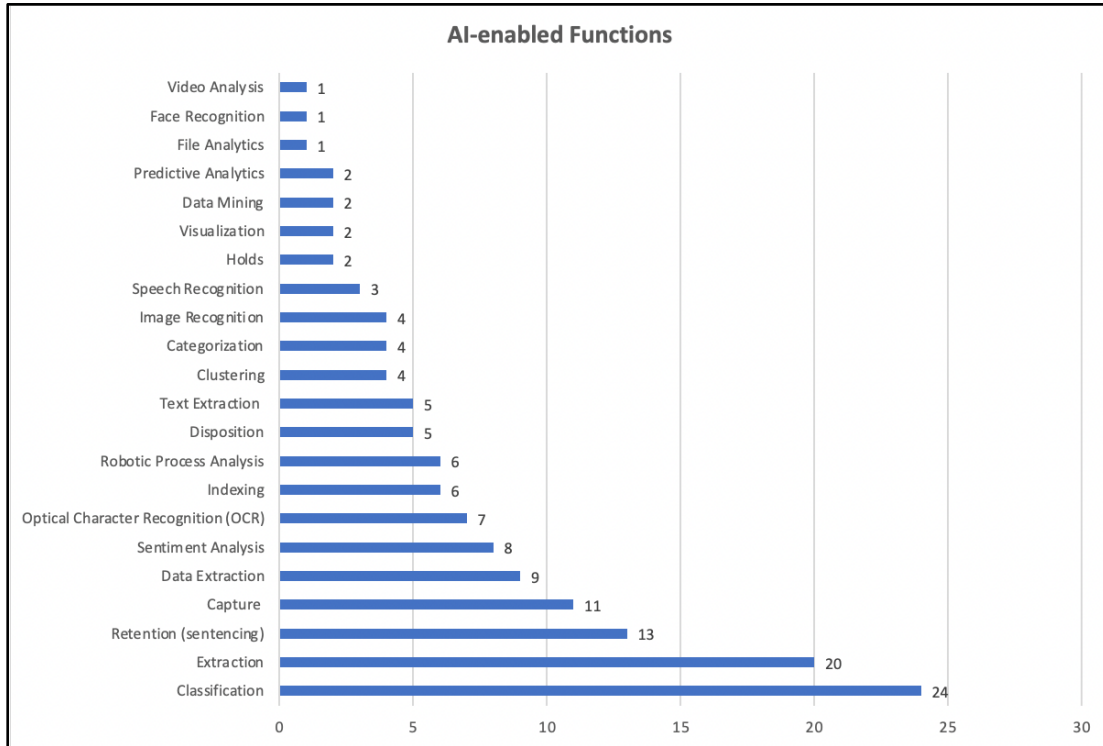**Figure 3.** Vendor materials reference content, document, record, and email.

Of interest to recordkeepers may be the fact that of the 19 products/services that include the term record, 4 are from the same vendor, Hyland. Among the remaining 15 are familiar names to records managers including IBM, Laserfiche, Microsoft, and OpenText.

**Functions of AI processes**

By far the products and services investigated appear to use AI for classification and extraction (not specifically detailing what type of extraction).  A positive finding is that 13 of the 25 reviews include the term retention but only 5 refer to disposition. In today's regulatory environment, disposition including the acts of destruction and redaction are becoming

increasingly necessary for compliance and risk management. This is an area that would benefit from AI technologies and tools. AI-enabled features cited in publicly available materials are shown in Figure 4.



**AI-enabled Functions**

| Function | Value |
| --- | --- |
| Video Analysis | 1 |
| Face Recognition | 1 |
| File Analytics | 1 |
| Predictive Analytics | 2 |
| Data Mining | 2 |
| Visualization | 2 |
| Holds | 2 |
| Speech Recognition | 3 |
| Image Recognition | 4 |
| Categorization | 4 |
| Clustering | 4 |
| Text Extraction | 5 |
| Disposition | 5 |
| Robotic Process Analysis | 6 |
| Indexing | 6 |
| Optical Character Recognition (OCR) | 7 |
| Sentiment Analysis | 8 |
| Data Extraction | 9 |
| Capture | 11 |
| Retention (sentencing) | 13 |
| Extraction | 20 |
| Classification | 24 |

**Figure 4.** AI-enabled Functions as observed by researchers reviewing publicly available Vendor materials.

**Limitations of data gathering:** While a standard data gathering template was used, the data was gathered by several different individuals who may have spent different amounts of time on the review and interpreted the materials differently. The data was gathered primarily through publicly available materials mainly available on the vendor site, which would present only positive descriptions of products/services. The vendors did not have an opportunity to review and correct our findings, which means there may be errors in the forms used to describe each product/service. For this reason, we caution the reader to use this material as a place to begin to better understand the terms and technologies described by various vendors when discussing management of content, documents, and records. The next phase of the study will include interviews with 25 of the vendors listed in the report about their products and services.