

## Datasheet

# mosaic agnitio

One of the fastest Intelligent Document Processing Platforms, Agnitio is a versatile AI-driven unstructured data processing platform for intelligent business process automation and knowledge mining.

[#BusinessProcessAutomation](#) [#UnstructuredData](#) [#DeepLearning](#)

Today, businesses and organizations find themselves overwhelmed by the huge volumes of data often in unstructured formats such as image, PDF, and Doc that they generate. Analyzing and processing such data manually is laborious and prone to errors. Further, the complexity and variety of formats make information extraction further difficult and expensive.

LTI's Mosaic Agnitio is a data extraction framework that offers cognitive services to extract data from documents in PDF, DOC, DOCX, XLS, and XLSX formats, as well as images in TIFF, JPEG and PNG formats. Tabular data from documents and images can also be extracted in structured formats such as CSVs and JSONs.

Mosaic Agnitio leverages Artificial Intelligence and Deep Learning to process the semi/unstructured data which is generally inaccessible and not easy to process. Mosaic Agnitio facilitates building data extraction solutions contextual to a business domain by leveraging specialized AI models. The platform encourages smarter business process automation and complements conventional RPA driven approach.

## Abilities Enabling Intelligent Document Processing and Knowledge Mining

### Entity / Knowledge Extraction

Data can be extracted from PDF, DOC, DOCX, XLS, and XLSX Images - TIFF, JPEG, and PNG

### Tabular Data Extraction

Extract tabular data from documents as What You See Is What You Get

### Solution Blueprinting

Solution for document processing can be setup using cognitive and other processing functions

### Sandbox to Production

All in one 'Solution' based approach for designing to training to execution

### Built-in AI Model Tutoring

Tutoring interfaces for ensemble of cognitive services for business to impart knowledge

### Open Integration with Upstream and Downstream

Data Ingestion from a variety of sources and API integration with downstream systems available

### Multiple Language Detection

Eases process automation for Global MNCs by recognizing and extracting data from 20+ Languages (English, French, Spanish etc.)

## Product Features



### Cognitive Information Mining Framework

Mosaic Agnitio is built on a foundation of Cognitive Information Mining Framework which offers versatility to handle variety of unstructured data. Information Mining solutions can be set up easily to process documents and images at scale.

<b>Solution Configuration</b>	Configure the solution with ensemble of cognitive services and extraction mechanism, train data extraction services and launch the solution for users.
<b>OCR Integration</b>	Images can also be processed easily by integrating with on-premise or on-cloud OCR engines. Mosaic Agnitio provides plug-and-play integration with best in class OCR Engines.
<b>Cognitive Processing</b>	Ensemble of models trained can be executed on a variety of documents for document classification, entity extraction, tables extraction, knowledge nugget extraction such as paragraphs, embedded images etc., topics and sentiment analysis.
<b>Solution Execution</b>	Process the documents in batch or in an interactive mode. Expose APIs on the solution for integration with RPA or with custom built user interface.
<b>Data Connectors</b>	Connectors to enterprise data sources such as BLOB, GCS, File systems etc. available with the framework to ingest unstructured data.
<b>Model management</b>	AI Models across the data extraction solutions on Agnitio can be scaled for build, deploy, retrain and execute at enterprise scale.
<b>Data Export</b>	Extracted data can be downloaded in JSON or CSV format for downstream consumption. Tabular data can be extracted in bulk in CSV format.



## Cognitive Services for Data Extraction

A suite of Artificial Intelligence services to help extract data from semi/unstructured documents. User can impart their business knowledge to train machine to process the documents and images

Name Entity Recognition	Extract text from data elements such as paragraph, key-value pairs, data available in header-footer etc.
Classification (Documents/Image/Content)	AI-based classification service to classify documents, images, or content. Once the classification model is trained, it can be used in the data extraction workflow. Classification can be trained on bulk documents, with minimum interactions.
Object Identification	Documents and Images with complex data structures and templates need zonal OCR. Mosaic Agnitio provides object identification AI service to identify the 'areas of interests' with Object Identification AI service.
Knowledge Nuggets Extraction	Cognitive engines to identify and extract sections in the document such as sentences, paragraphs, headings, embedded images and tables with their captions, URLs etc. Further analyzing the topics, sentiments and summary of the given text.



## Train as you Go

Train machine on different cognitive services such as document classification and name entity recognition for adaptive extraction easily without the involvement of data scientists

AI Tutor	Label the data to train the machine on various cognitive services through user-interface. An ensemble of AI models can be trained with ease.
Training/ Re-training	Train and Tune AI models iteratively using user interface – foolproof way which does not require data scientist's knowhow.
Feedback	Provide feedback to AI models by labeling or deleting predicted entities on the job while processing the documents.
Continuous Learning	Continuous learning on the job on the user feedback with automated model deployments and notifications. Data extraction accuracy increases as more and more documents are processed.



## **Integrated Verification Station**

Integrated verification function that can be embedded in any data extraction solution built using the framework.

### **Document Layout**

QC forms can be personalized using document layout configurations. Provision to setup the QC form with entities, repeating sections in the document (groups and nested sub-groups) to make the QC process intuitive.

### **Validations**

Automated validation setup that highlights discrepancies pro-actively in the QC form. Validations such as mandatory values, drop-downs, range of values, numeric, and alpha-numeric can be configured.

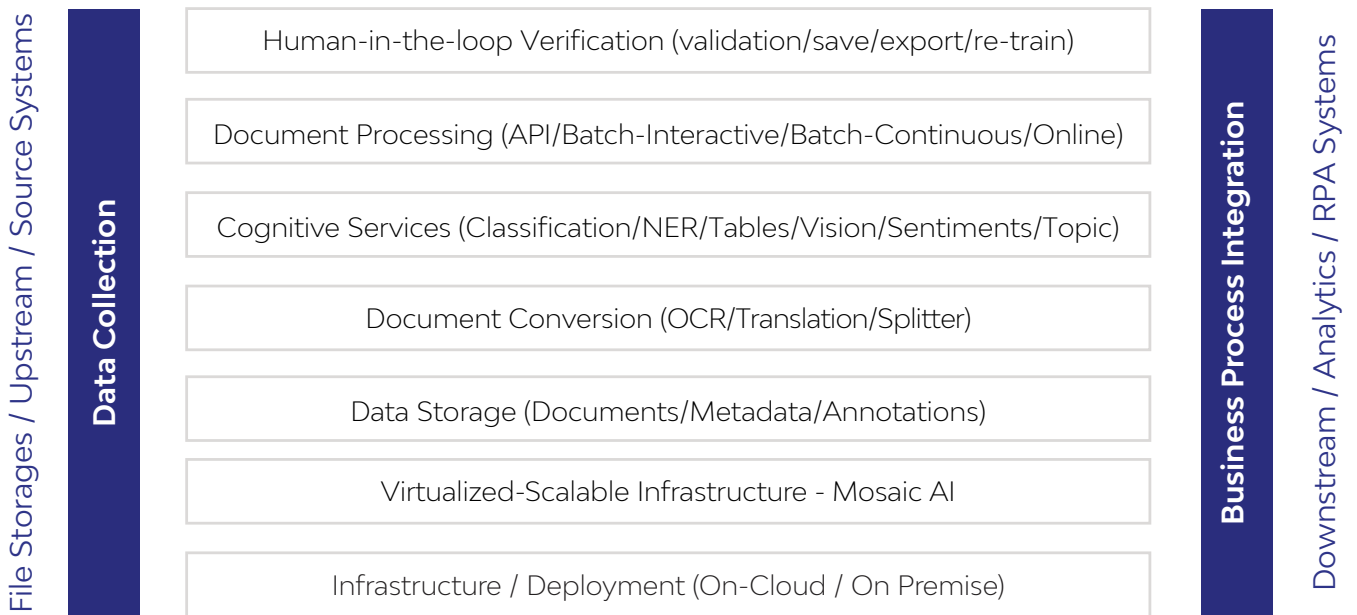
### **Edit-in-place**

Enrich the data by edit-in-place for downstream consumption.

### **Intuitive QC Form**

Integrated QC to visualize your document as it passes through the various stages of processing: original document, converted text, extracted entities are all shown side by side for traceability.

## Product Architecture -



## Deployment Options

Flexibly deploy on cloud or on-premises

