



# Audax Labs – DocuLens OCR



# Our Capabilities & Experience

Audax labs is an Innovation Partner with a strong System Integrator background. We work with enterprise clients in their innovation journey from ideation to enterprise grade deployment.

## PARTNERS

**HITACHI**  
Inspire the Next

**Microsoft**  
Solutions Partner

## CUSTOMERS

**HITACHI**  
Inspire the Next

**Microsoft**

**Parker**



**LiveWire**

**Rabobank**

**9to5**  
seating

**Kelvin**






**Bonsai**  
mediagroup



## INDUSTRIES

-  Automotive
-  Manufacturing
-  Healthcare
-  BFSI
-  Retail

## TECHNOLOGIES

-  **AI**  
Artificial Intelligence
-  **AR, VR, & XR**  
Augmented Reality
-  **IoT**  
Internet of Things
-  **Cloud**  
Storage & Computing
-  **Data**  
Managing Data lifecycle

## GLOBAL PRESENCE



# Organization Data Challenges

The world of business is data-focused. Organizations are continuously assessing and adopting new technologies, and information management practices. With the IT landscape continuing to evolve, it gets even more difficult to ensure that the right data is at the right place, at the right time.

Data is constantly being generated through diverse platforms and the source of data for any organisation is no longer exclusive to traditional applications. Implementing strategies to manage data and drive business requirements is a challenging task and often there are no fixed remedies to handle data related issues.

Some common challenges which enterprises face today are:



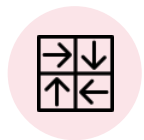
Poor Data Quality



Data Security



Data Duplication



Disparate Data Sources



Data Discovery



Data Analysis



Exponential Data Growth



Cost of Managing Data




Data Compliance

# DocuLens OCR by Audax Labs

Audax Labs DocuLensOCR Engine is an enterprise OCR solution that reads document residing in customer's modern cloud platforms like Azure Storage or legacy storage systems such as CIFS, NFS, or Hitachi Content Platform. OCR Engine supports all standard image and document formats, such as PDF, JPG, PNG, BMP, TIFF, and GIF. For PDFs, the OCR Engine breaks each page into individual images (such as TIFF) and then runs the text extraction process. DocuLensOCR solution includes system management unit (SMU) UI interface for managing cluster/worker instances.



## COST EFFECTIVE

 OCR helps in reducing several costs like printing, copying, shipping charge, hiring manpower for data extraction etc.

## IMPROVED EFFICIENCY



Benefit Of OCR is its ability to automate the process of data entry and text recognition, saving time and reducing risk of errors.

## SEARCHABLE DOCUMENT



It is easy to make a document searchable by using Keywords and Phrases Contained In The Document.

## BETTER ACCESS TO INFORMATION



Benefit Of OCR Is Its Ability To Digitize Paper Documents/Images , Making Them Easier To Store , Search & Access.

# Features of DocuLensOCR Services

Audax Labs Document Engine converts and extracts the images from the input file (PDF, JPG, PNG, Doc etc.) and these images are further converted to a file containing the text.



DocuLensOCR offers central system management unit (UI) Services to manage processing, monitoring and reporting in real-time.

Central  
SMU



It allows document conversion into searchable TEXT/XML allowing faster searches and index.

Efficient  
Processing



Robust AI and ML with other technologies detect the logical structure of a document and it can further be used to make smart decisions.

Powered with  
AI & ML



Using multi-core CPUs and processing images in parallel on multiple threads, the OCR steps can be performed significantly faster.

Rapid  
Processing



It can be deployed on-premise, virtual environments or Cloud platforms like Microsoft Azure or Amazon Web Services.

Compatibility &  
Support



De-skewing, rotation, distortion correction, text line straightening, page splitting, adaptive binarization, ISO noise reduction and other image correction steps deliver best OCR results.

Image  
pre-processing



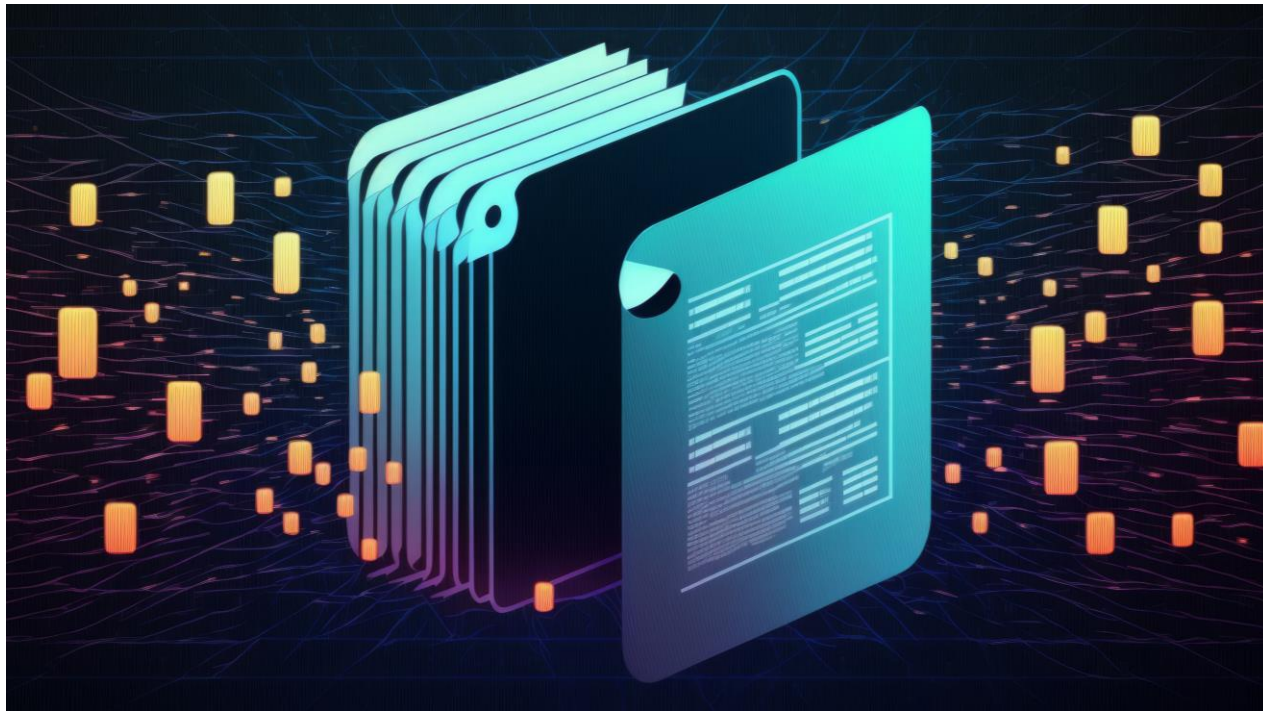
# Multilingual Support

Audax Labs' OCR solution comes with multi-language support. Currently, it can process and generate output for 120+ languages.

- ✓ Arabic
- ✓ Czech
- ✓ Danish
- ✓ German
- ✓ English
- ✓ Spanish
- ✓ Estonian
- ✓ Persian
- ✓ Finnish
- ✓ French
- ✓ Hindi
- ✓ Hungarian
- ✓ Indonesian
- ✓ Italian
- ✓ Japanese
- ✓ Korean
- ✓ Lithuanian
- ✓ Macedonian
- ✓ Dutch; Flemish
- ✓ Norwegian
- ✓ Polish
- ✓ Portuguese
- ✓ Russian
- ✓ Slovak
- ✓ Slovenian
- ✓ Albanian
- ✓ Swedish
- ✓ Turkish
- ✓ Ukrainian



# Dark Data Risk Mitigation with the help of DocuLens OCR



## Challenge:

A large Wealth Management company seeks a robust solution to scan and analyze over 40 million documents, comprising largely unstructured dark data, with the aim of identifying sensitive information within them. The organization faces significant compliance and security risks due to the uncertainty surrounding this dark data, necessitating an efficient and accurate method for content analysis and risk mitigation.

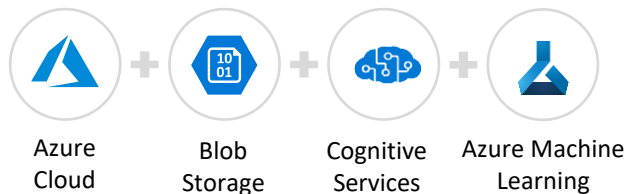


## Solution:

Audax Labs develops DoculensOCR, a cutting-edge solution, harnessing the power of Machine Learning and Computer Vision technologies. This innovative solution is specifically designed to address the challenge faced by the customer. This solution is capable of identifying sensitive information from multiple formats of image files and PDF document. OCR reduce regulatory risks as it processes scanned images and pdf files to text. Any Customer indexing Solution can be integrated with OCR solution to consume data from OCR results.



## Solution Component:



## Outcome:

- Data Discovery and Compliance
- Data Security
- Reduced operational cost
- Efficiency and accuracy



# Outcome Driven Innovation!

---