DocStudio

AI-Driven Digital Document Framework:

Transforming and Optimizing Supply Chain Processes

Speaker:

Alisa Konchenko, VP of Business Development

2025 PIDX Spring Conference

Agenda

Challenges of Manual Document Management

Current options for document recognition and item matching: Common document types for OCR and AI automation

Modern approach offered by eDMS

Document automation case studies

Breakdown of costs and benefits offered by AI-automation

Disadvantages of Manual Document Management

Reduced efficiency and productivity

Difficulties in tracking documents and retrieving their data
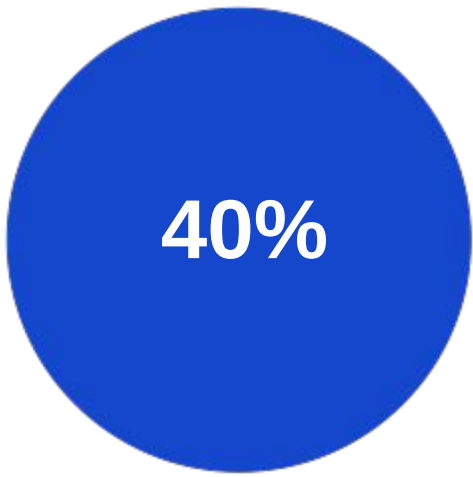
Poor document security and compliance risks

Risk of document loss or damage
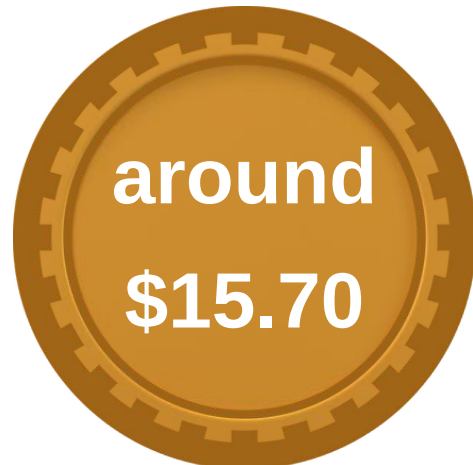
Higher administrative and labor costs

Slower response times to business needs

Why Go Digital

40%

of document-related costs can be reduced when using a DMS

up to 80%

of time spent file-sharing can be saved when using a DMS

around $15.70

per transaction is spent on average for exchange of paper documents

45%
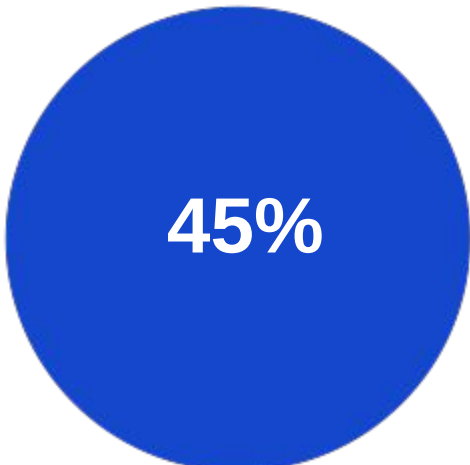
of companies spent at least 1 week to have a contract signed

Enterprise Document Flows in Supply Chain

Supply Chain Documents

Invoices

Purchase Orders

Delivery Orders

Inventory Reports

Goods Receipts

Advanced Shipping Notices

Requests for Quotation (RFQ)

Bills of Lading
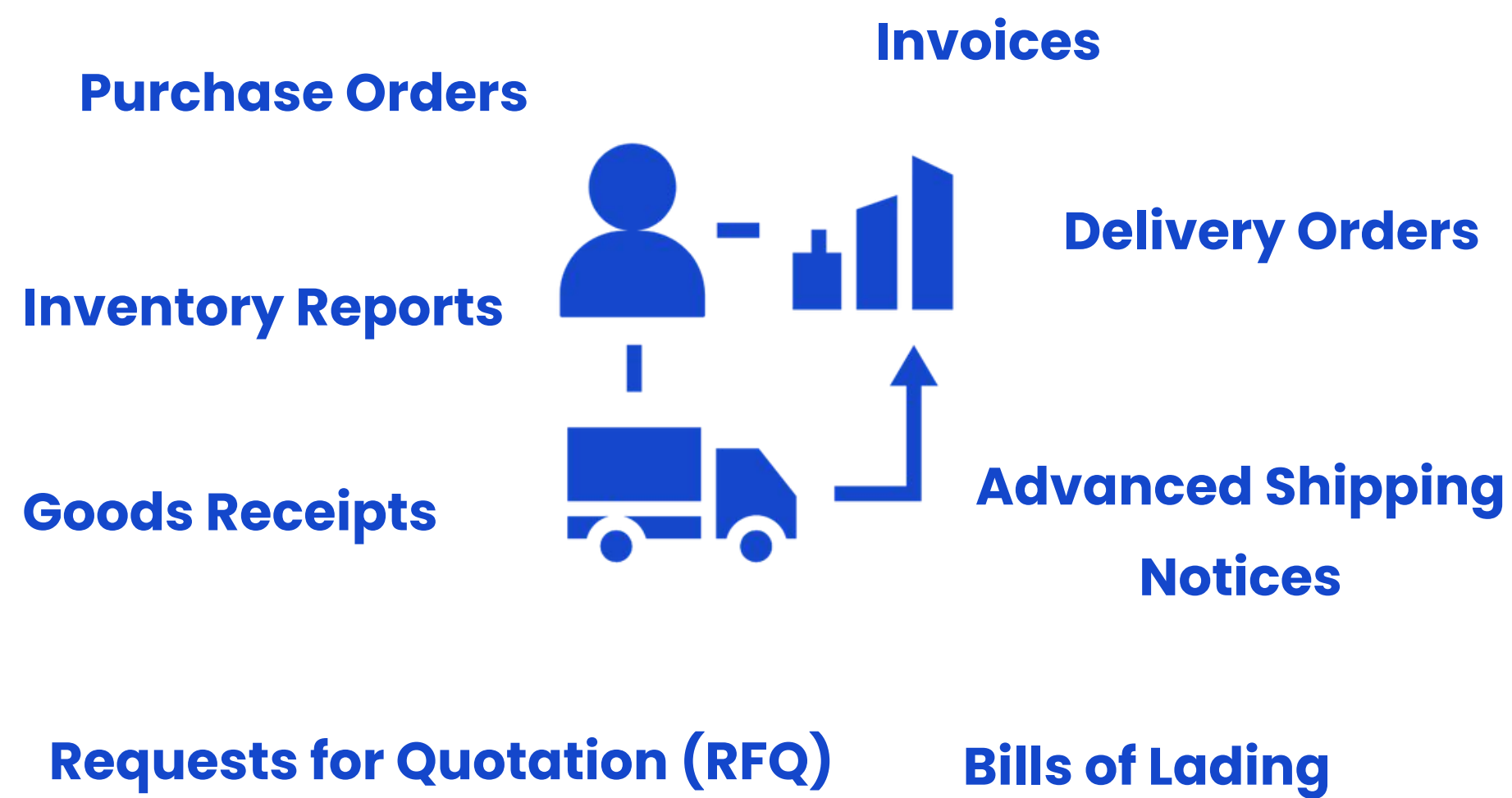
Legal and Operational Documents

Employment Agreements

Contracts

Partnership Agreements

NDAs

Terms and Conditions

Compliance Forms

Service Level Agreements
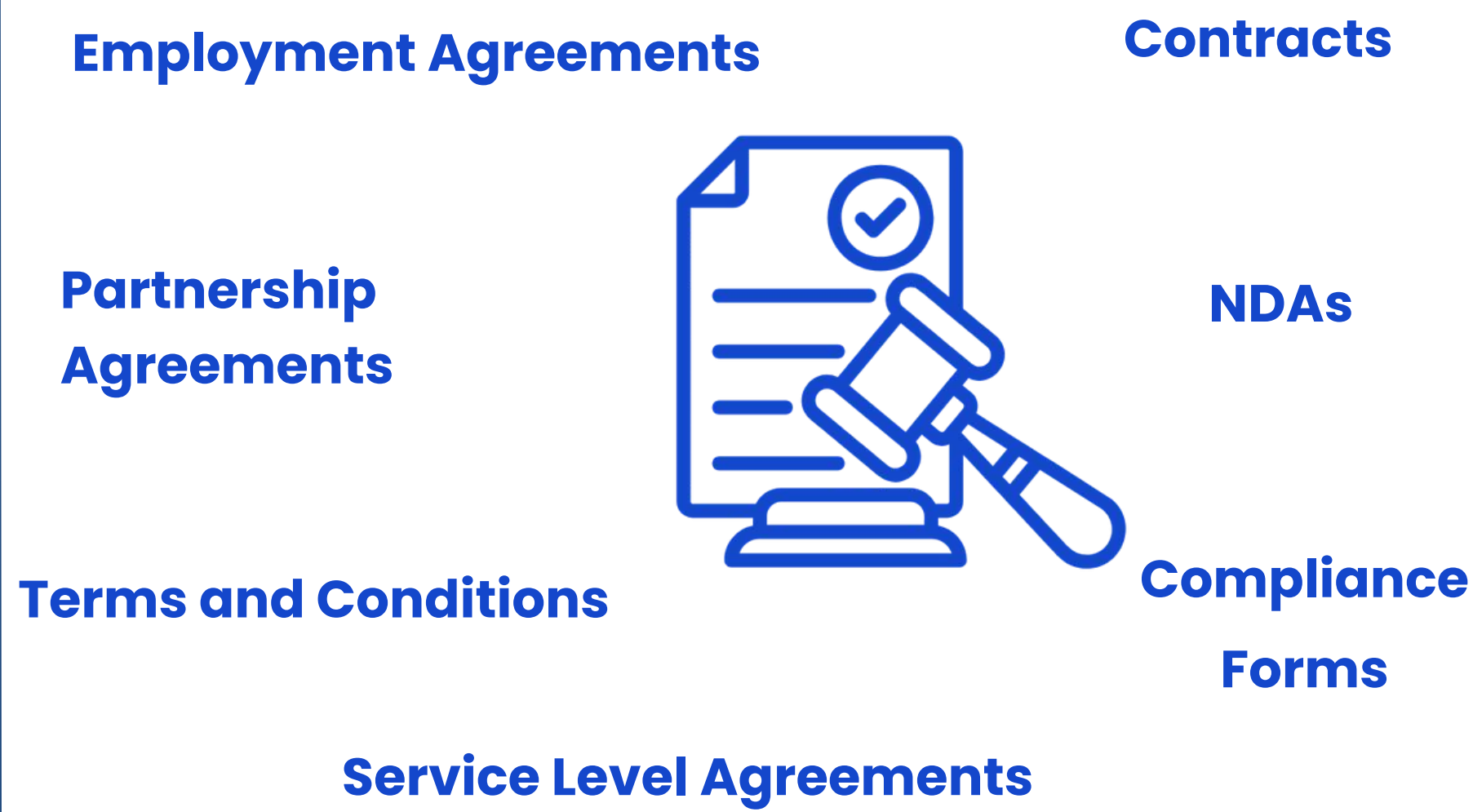
Supply Chain Documents

Challenge: lack of master data and incomplete EDI integration /automatization

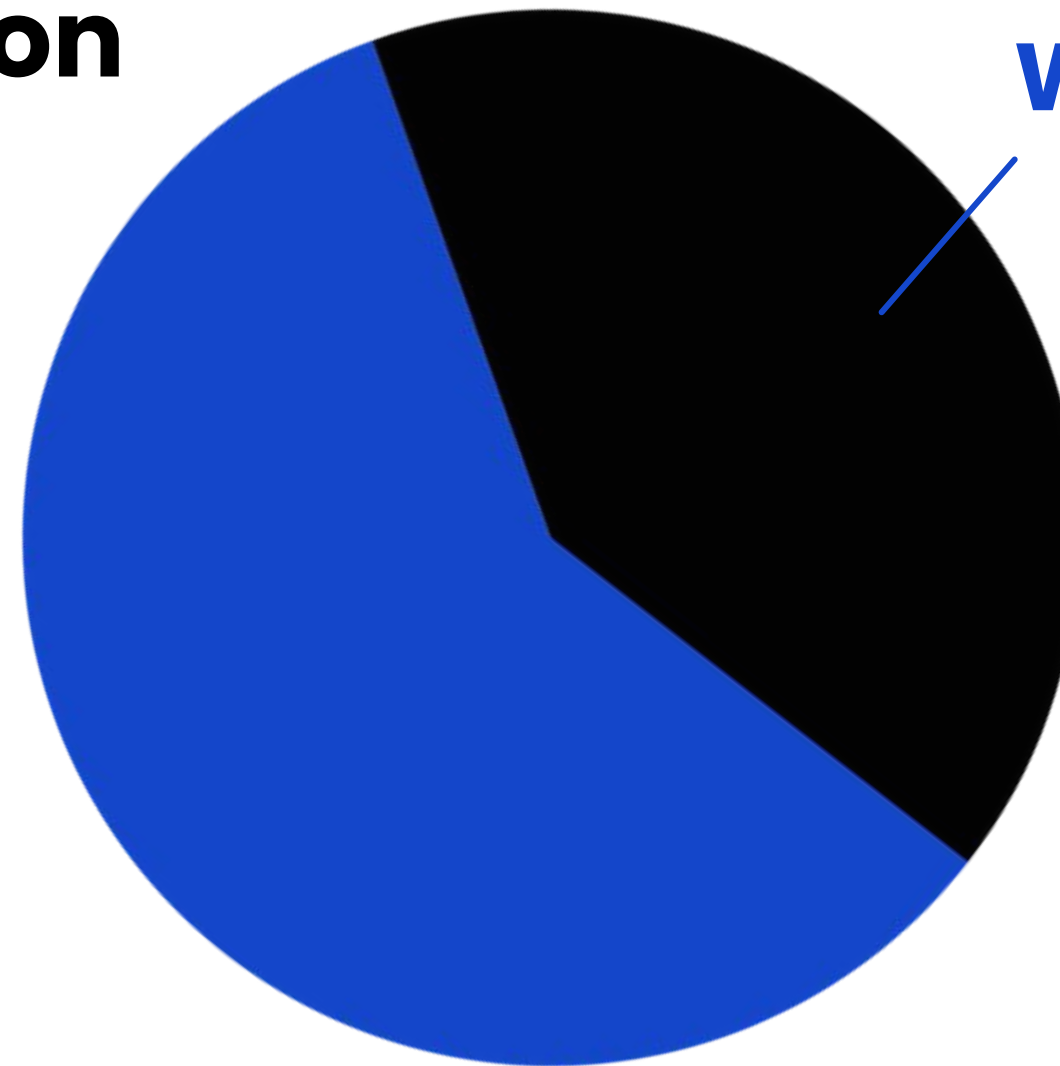**59%** of US companies in Supply Chain adopted EDI for their daily operations

What about the other 41%?

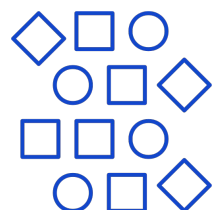Source: Data Interchange. (2023, October 5). *Over 41% of companies at risk without EDI.* https://datainterchange.com/over-41-of-companies-at-risk-without-edi/

While EDI systems work smoothly, not all your partners will have the necessary infrastructure to use them.
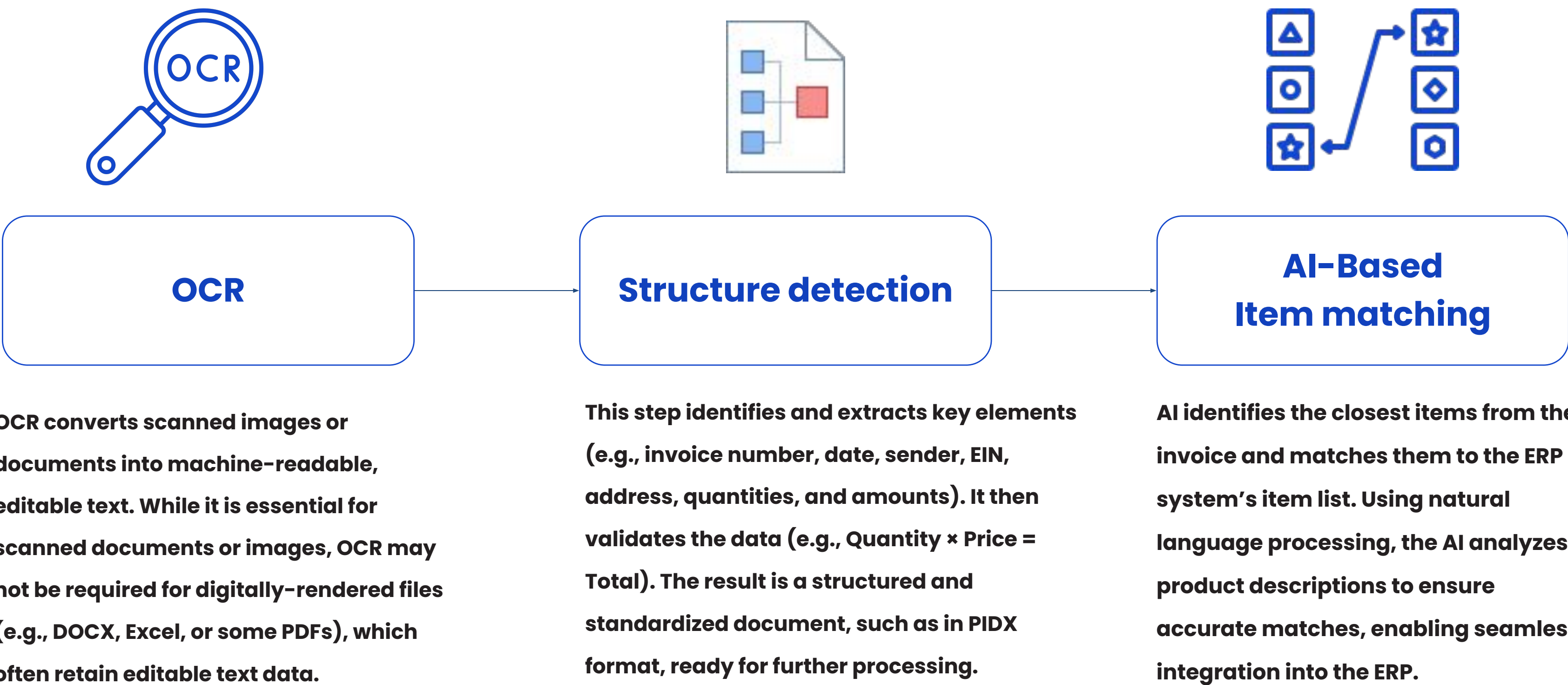
Main Challenges with Invoice Recognition

- **Complex Layouts:** Some invoices are difficult to recognize due to complicated designs and layouts.
- **Data Confusion:** Invoice numbers may be mixed up with dates, and columns like Quantity, Price, and Total Amount can be misinterpreted.
- **Handwritten Marks:** Handwritten marks are inconsistently processed. Sometimes they should be ignored, but in other cases, they need to be considered.
- **Decimal and Cents Processing:** Decimal separators and ".00" in cents require accurate processing to avoid errors.
- **Unstructured Data:** Even after successful OCR (Optical Character Recognition), the extracted data may still lack structure.
- **Manual Item Matching:** Item matching against the ERP item list still requires manual intervention.

Finding Solutions: Three Steps to Take

OCR

OCR converts scanned images or documents into machine-readable, editable text. While it is essential for scanned documents or images, OCR may not be required for digitally-rendered files (e.g., DOCX, Excel, or some PDFs), which often retain editable text data.

Structure detection

This step identifies and extracts key elements (e.g., invoice number, date, sender, EIN, address, quantities, and amounts). It then validates the data (e.g., Quantity × Price = Total). The result is a structured and standardized document, such as in PIDX format, ready for further processing.

AI-Based Item matching

AI identifies the closest items from the invoice and matches them to the ERP system's item list. Using natural language processing, the AI analyzes product descriptions to ensure accurate matches, enabling seamless integration into the ERP.

Results must be human-approved to ensure the quality of final data and can then be imported to the ERP.

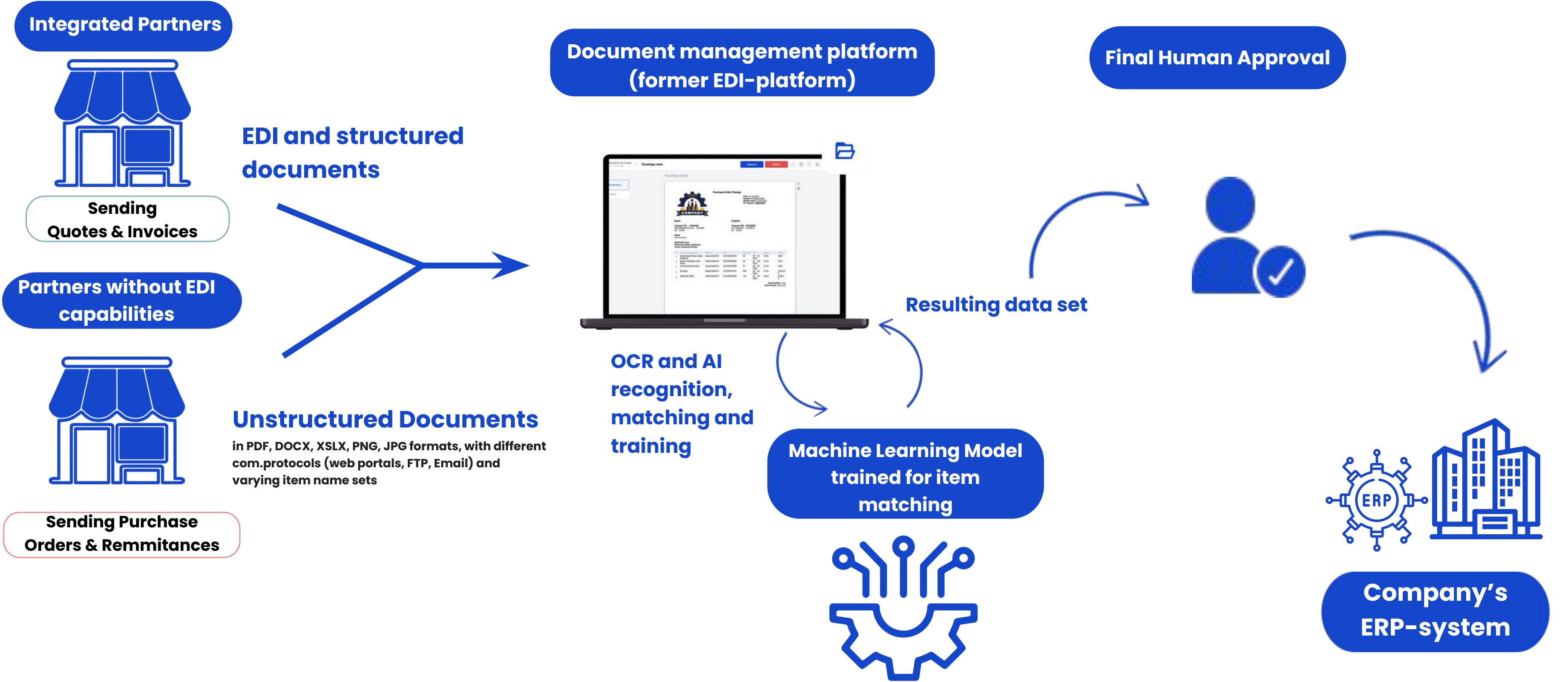# Available Software for Document Recognition: OCR and AI Automation Options

There are some products and SaaS on the market providing document structure recognition:

Several products and SaaS platforms offer document structure recognition, available through web platforms, API-only options, or iOS/Android applications.

However, these solutions typically lack advanced item matching capabilities and do not offer integration with existing EDI (Electronic Data Interchange) or DMS (Document Management Systems).

AI Document Integration Framework

Integrated Partners

Sending Quotes & Invoices

EDI and structured documents

Partners without EDI capabilities

Sending Purchase Orders & Remmitances

Unstructured Documents

in PDF, DOCX, XSLX, PNG, JPG formats, with different com.protocols (web portals, FTP, Email) and varying item name sets

Document management platform (former EDI-platform)

OCR and AI recognition, matching and training

Machine Learning Model trained for item matching

Resulting data set

Final Human Approval

Company's ERP-system

AI Document Integration Framework Description

Ongoing Daily Processes:

Document Ingestion

Hundreds or thousands of predefined documents (e.g., quotes, invoices, orders, specs) arrive through various channels (web portal, procurement platform, FTP, email) and formats (PDF, DOCX, PNG, JPEG).
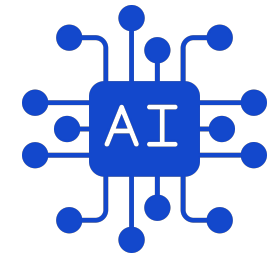
OCR Processing

Image files are processed with OCR to extract machine-readable text content.

Block Detection and Decomposition

Documents undergo block detection to extract key elements (e.g., DocumentType = "Invoice", DocumentDate = "November 15, 2024", SenderEIN = 352589633).

Data Storage

Detected data is stored in a structured format (e.g., PIDX). **At this point, items are not yet matched with the ERP system.**

AI-Based Item Matching

AI matches items in the document with corresponding items in the sender's ERP item list, improving data accuracy.

Approval Process

Invoices are sent for approval by a designated person. Correct matches enhance system reliability, while incorrect matches are used to retrain the AI model for future improvements.
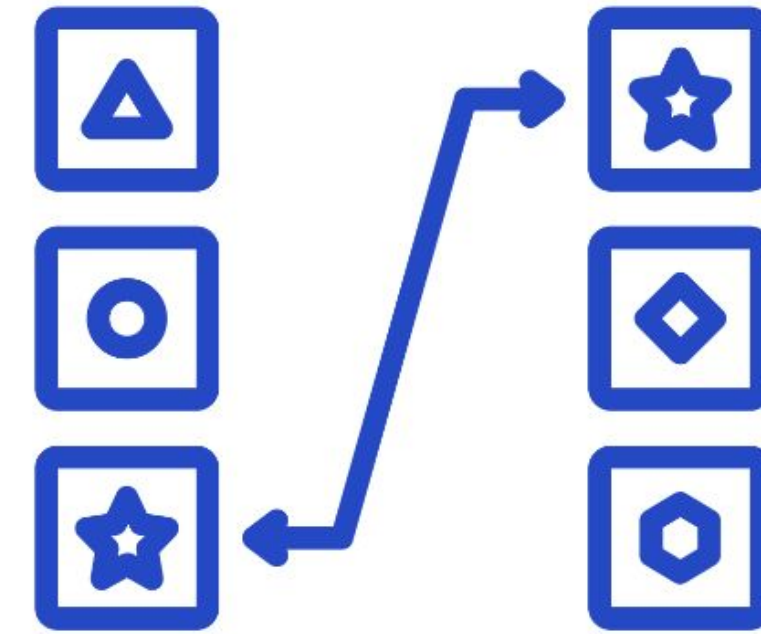
ERP/Other System Integration

Recognized and matched invoices, in predefined formats, are sent to the enterprise ERP system as regular EDI documents.

AI Document Integration Framework Setup Process

There are two AI-models to be used for such framework:

Document Structure Detection Model

Item Matching Model

At the initial stage, and whenever there are significant changes (e.g., document type set, design/layout updates, or major changes to the item list), the accuracy of the AI models should be reevaluated.

If needed, the models should be retrained using updated data. While this process may be ongoing, it requires a supervisor to act as a quality "judge" to ensure the system's accuracy and reliability.

Unstructured Documents

AI Models
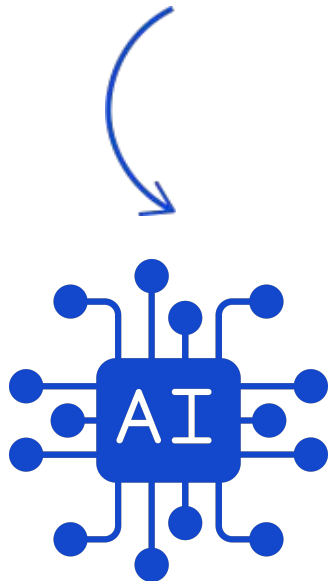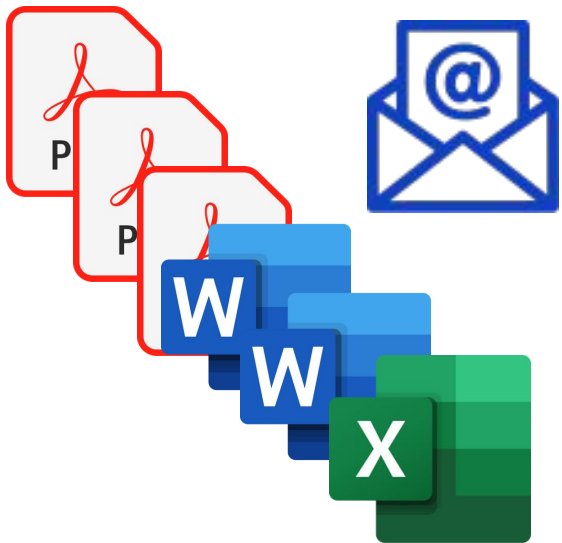
Structured Documents

Purchase Order Change

Date: 07/18/2024
Number: 873489743897
Delivery Date: 07/26/2024
PO Number: 326562356

COMPANY

Buyer:

Company TGF , 64646464
2543 Weddington Ave. , Charlotte
NC , 87654

Supplier:

Company ABC , 930432081
167 Atlanta St , ATLANTA
GA , 30755

Store:
Store Number

Skid/Pallet Type:
Receiving Facility Limitations:
Carrier Assigned by Buyer:

#	Product Description	Brand	UPC	Quantity	Unit	Price	Total
1	Rectangular Pillow, Ocean & Natural	Sweet BatterFly	657643567876	65	H3 : 18-Pack	56.00	3640
2	Black Crosshatch Latte Mugs	Sweet BatterFly	657643567865	45	4F : 100-Pack	67.00	3015
3	Aria Champagne Flute	Sweet BatterFly	657643567888	87	H3 : 18-Pack	67.00	5829
4	Big Bowl	Sweet BatterFly	657643567876	898	H6 : 30-Pack	32.56	29238.8 6
5	Deep Dish Plate	Sweet BatterFly	657643567898	123	H3 : 18-Pack	45.45	5590.3 5

Total Quantity: 1218
Total Amount: 47313.23

PIDX INTERNATIONAL

ERP System

SAP

From Unstructured to Structured Fast and Accurately

The resulting file is saved in both PDF and XML, with XML convertible to JSON, CSV, or EDI for seamless integration.

Before

After

ROI and Cost-Benefit Analysis

With an average company processing 500 transactional documents per month, or around 6000 documents annually.

Manual Handling

20 mins

spent processing one document
On average, it takes 15 minutes to create one document, with around 40-80 line items of information that need to be retyped manually in every transactional document.
An additional, 5 minutes are spent correcting errors and double-checking the information.

$8

in labor costs per document
With the average wage of 25$ per hour for employees working with document processes.

+

other operational costs
That includes the use of paper and ink, document storage, mailing fees, and others.

$~50,000 annually in labor and document processing costs

AI-automated processing

0.5 mins

spent processing one document
With ready-to-use templates and streamlined document workflows in place.

$0.6

cost per document processing
Including costs for labor and software license for 5 users.

$~9,780

annual total cost for the first year
Including software license for 5 users, project rollout, support and maintenance, manual labour costs involvement.
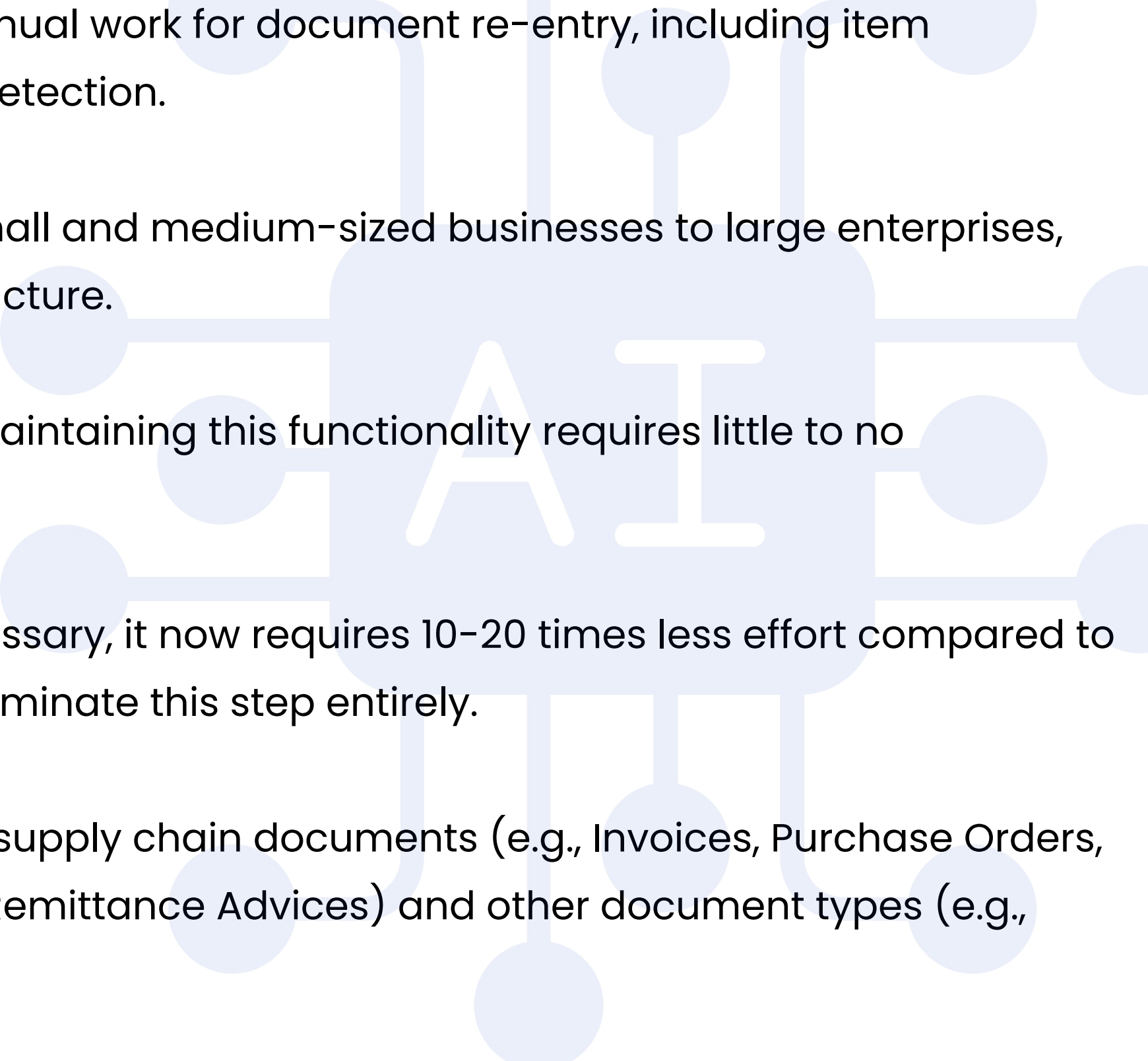
$~5,780

annual total cost for the second and subsequent years
Including software license for 5 users, support and maintenance, manual labour costs involvement.

ROI of ~ $40,220 (80.44%) in the first year of adoption and $44, 220 (88.44%) in the subsequent years

Takeaways

✔ **AI Automation:** AI can eliminate up to 90-95% of manual work for document re-entry, including item matching, document type detection, and structure detection.

✔ **Easy Adoption:** This innovation is accessible from small and medium-sized businesses to large enterprises, with minimal changes required to existing IT infrastructure.

✔ **Minimal AI Expertise Required:** Implementing and maintaining this functionality requires little to no specialized AI knowledge.

✔ **Human Approval:** While human approval is still necessary, it now requires 10-20 times less effort compared to manual processing. Future AI advancements may eliminate this step entirely.

✔ **Wide Application:** This approach is applicable to all supply chain documents (e.g., Invoices, Purchase Orders, Shipping Notices, Bill of Lading, Field Service Tickets, Remittance Advices) and other document types (e.g., agreements, contracts, application forms).

✔ **Role of Standards:** Standards and code lists play an even more crucial role as the "common language" for seamless integration and automation.

Q & A

Thank You for Your Time!

Alisa Konchenko, VP of Business Development
ak@docstudio.com