

# Intelligent Document and Unstructured Data Processing with Generative AI

**Ted Way, PhD**

Director of Product Management, SAP Business AI, SAP  
[ted.way@sap.com](mailto:ted.way@sap.com)

Las Vegas

---

**2024**

**SAP**insider

# Disclaimer

The information in this presentation is confidential and proprietary to SAP and may not be disclosed without the permission of SAP. Except for your obligation to protect confidential information, this presentation is not subject to your license agreement or any other service or subscription agreement with SAP. SAP has no obligation to pursue any course of business outlined in this presentation or any related document, or to develop or release any functionality mentioned therein.

This presentation, or any related document and SAP's strategy and possible future developments, products and or platforms directions and functionality are all subject to change and may be changed by SAP at any time for any reason without notice. The information in this presentation is not a commitment, promise or legal obligation to deliver any material, code or functionality. This presentation is provided without a warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. This presentation is for informational purposes and may not be incorporated into a contract. SAP assumes no responsibility for errors or omissions in this presentation, except if such damages were caused by SAP's intentional or gross negligence.

All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.



## In This Session

---

Explore the new premium features of Document Information Extraction

Gain insights into how generative AI is reshaping intelligent document processing and discover the key principles driving this revolution.

Delve into the intricacies of SAP Document Information Extraction.

Explore the extensive language support with over 40 languages in SAP Document Information Extraction.

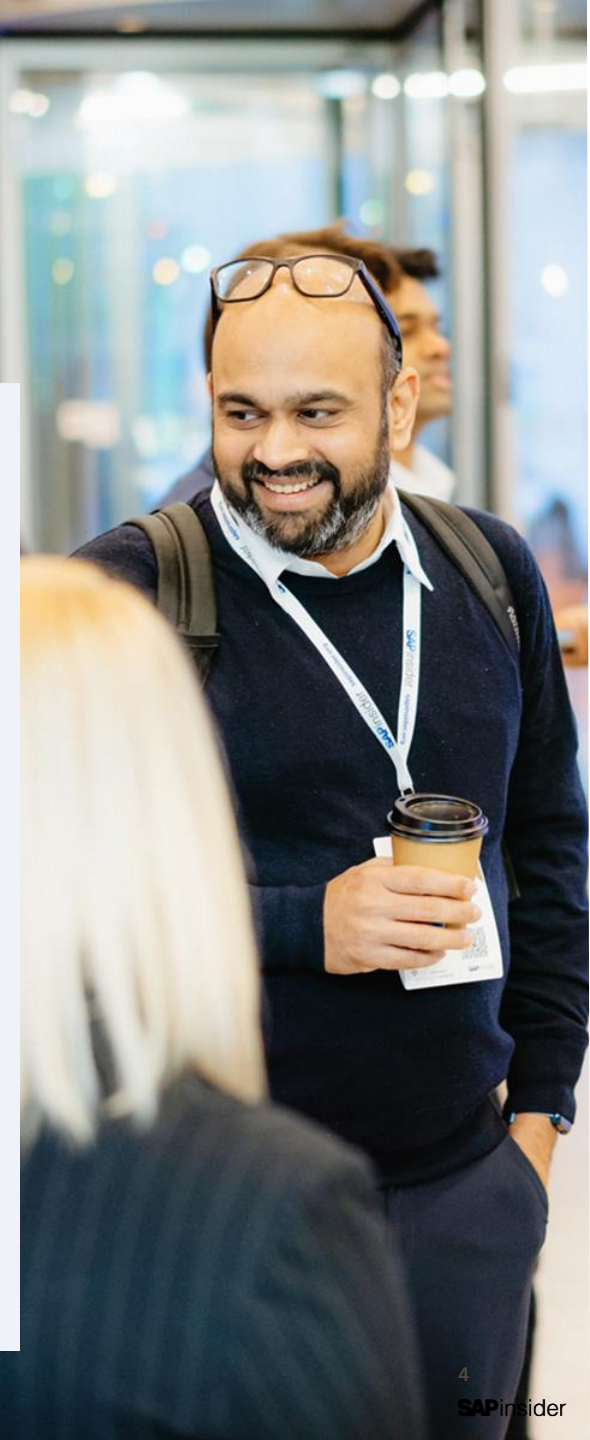
Get a preview of the roadmap for 2024



# What We'll Cover

---

- Document Information Extraction Premium Edition overview
- Demo
- Roadmap

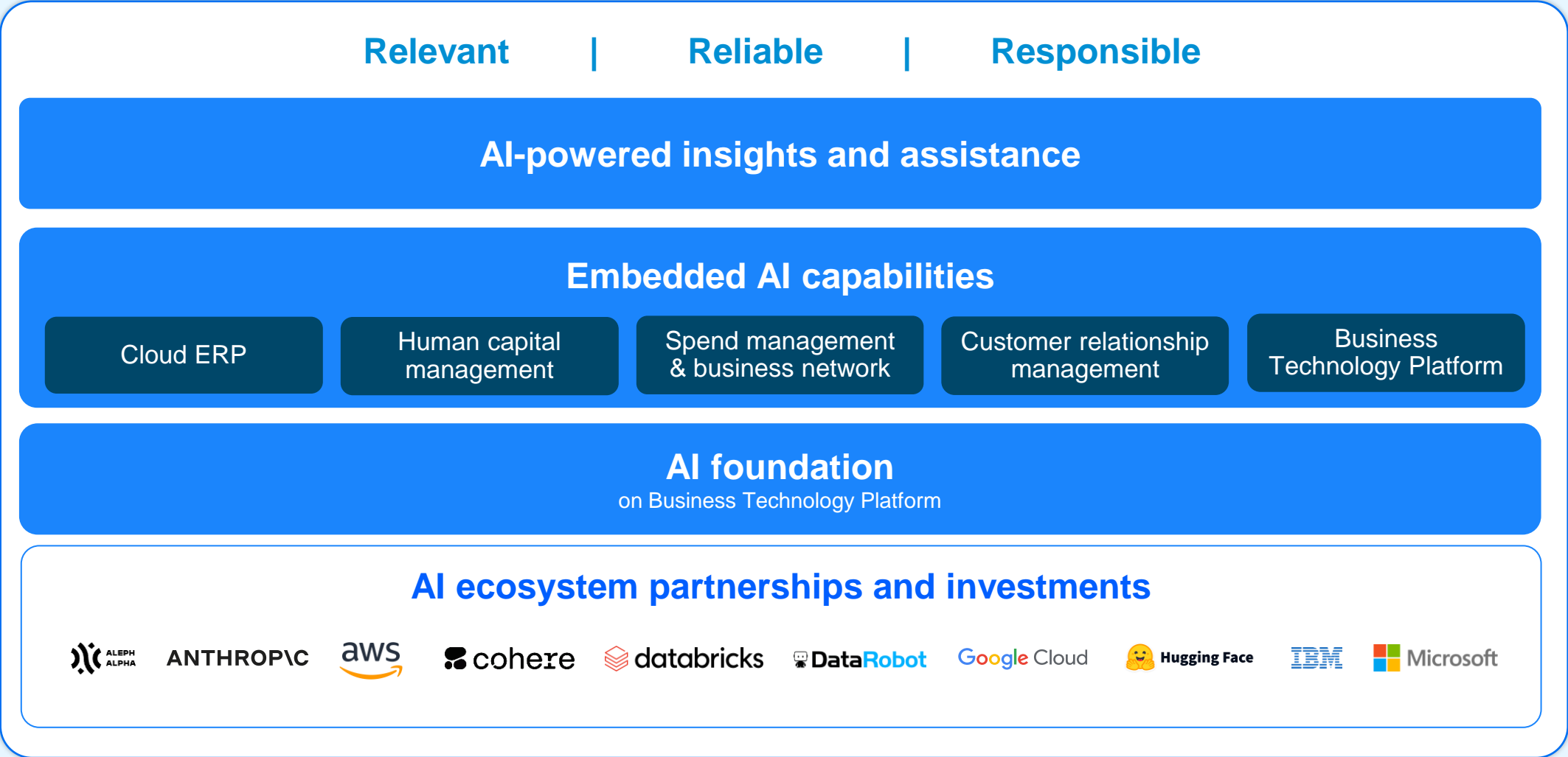


---

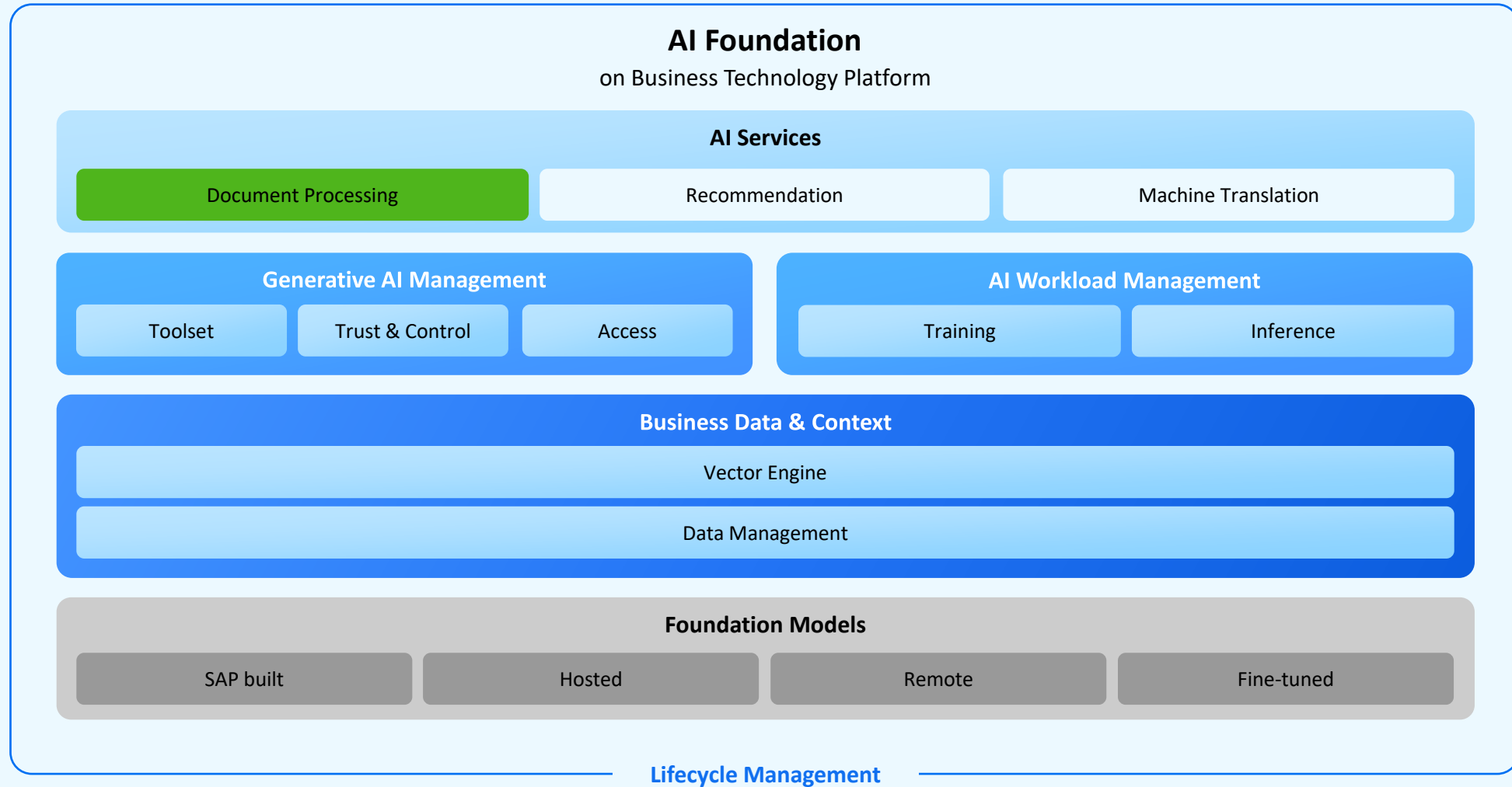
## **Topic 1: Motivation for document and unstructured data processing**



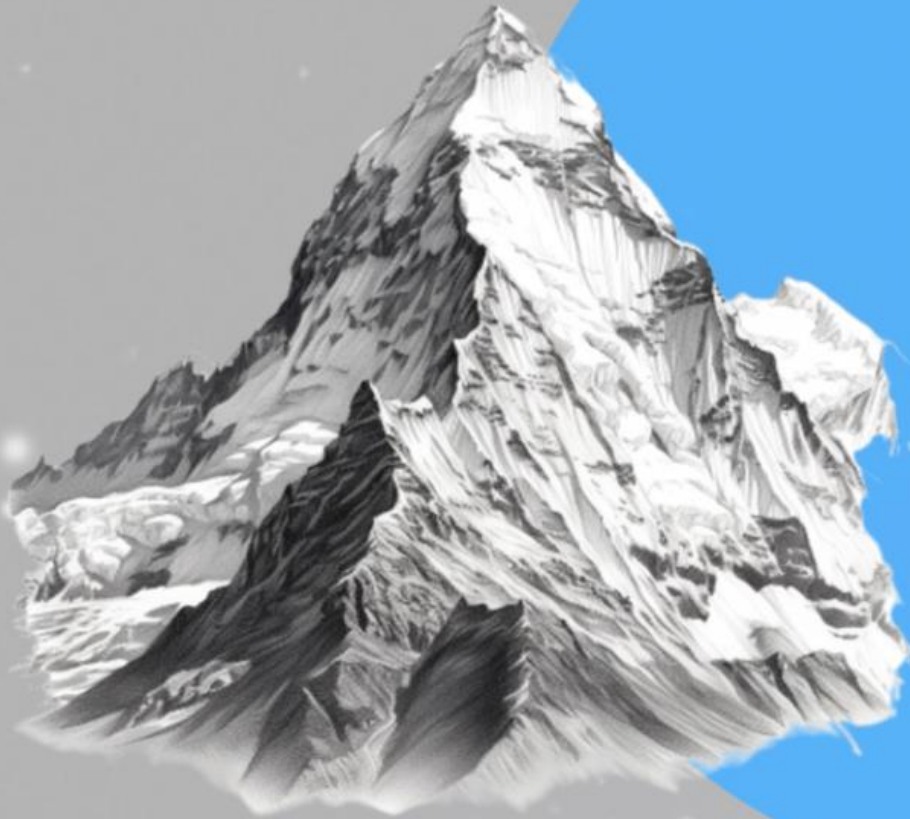
We believe in providing AI built for business. That is, AI that enriches all your business processes and empowers you to steer your business with agility.



# AI Foundation on SAP BTP



# AUTOMATED DOCUMENT PROCESSING



**Mount Everest**

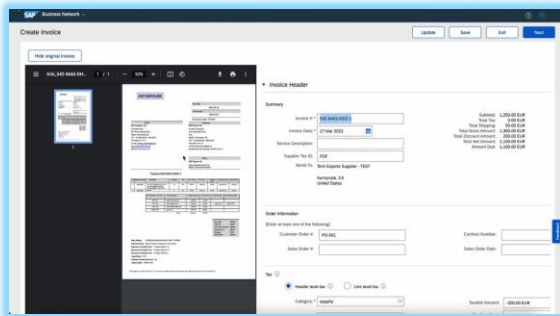


**Documents/Quarter**



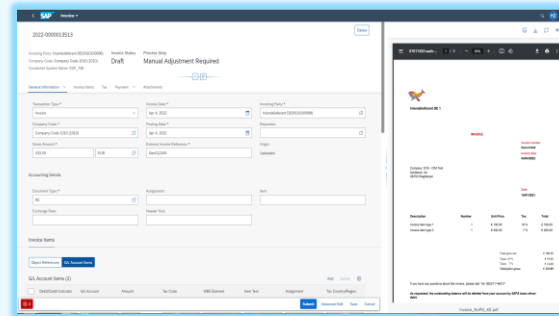
# Business process-specific AI for Document Processing in SAP

## SAP Business Network Intelligent Invoice Conversion



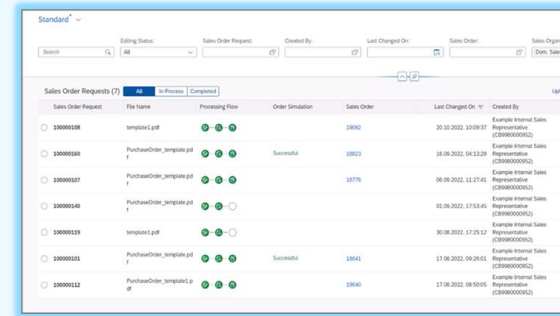
- Improve operational efficiency through simple invoice upload and automation for invoice conversion
- Increase invoice accuracy by reducing manual and error-prone entry
- Accelerate invoice payments by reducing the number rejected invoices by your customers
- See in action: [Intelligent Invoice Conversion](#)

## SAP Ariba Central Invoice Management



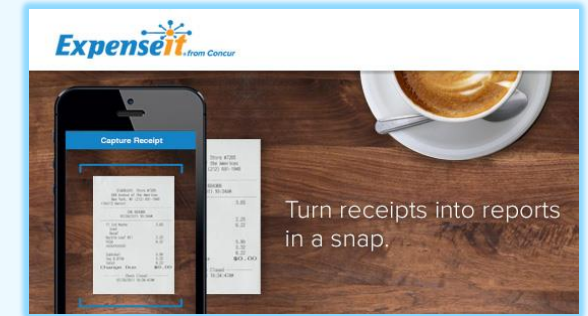
- Provides inbound handling via Business Network and central processing of invoices.
- Deployed on BTP available for SAP S/4HANA Cloud, public edition
- See in action: [Central Invoice Management](#)

## SAP S/4HANA Sales Order Automation



- Enables faster completion to prevent delays in the sales order delivery
- Reduces redundant tasks to free time for users
- Reduces the risk of human errors in the process
- See in action: [Sales Order Automation with AI](#)

## SAP Concur Expenselt



- Images of receipts are captured in the Expenselt app
- Information extraction starts on the device with ML models analyzing the image
- Improved employee productivity and overall audit efficiency
- See in action: [Expenselt](#)

Document Information Extraction

Responsible AI

# The Cost of Unstructured Data

Processing of unstructured data causes high costs, missed insights, and operational bottlenecks impacting growth and efficiency.



**17.60€**

AVERAGE COST PER INVOICE

- BILLEMENTIS

**80-90%**

OF ENTERPRISE DATA IS UNSTRUCTURED

- GARTNER

**4 trillion**

PAPER DOCUMENTS ANNUALLY IN THE US

- EMERGEN RESEARCH

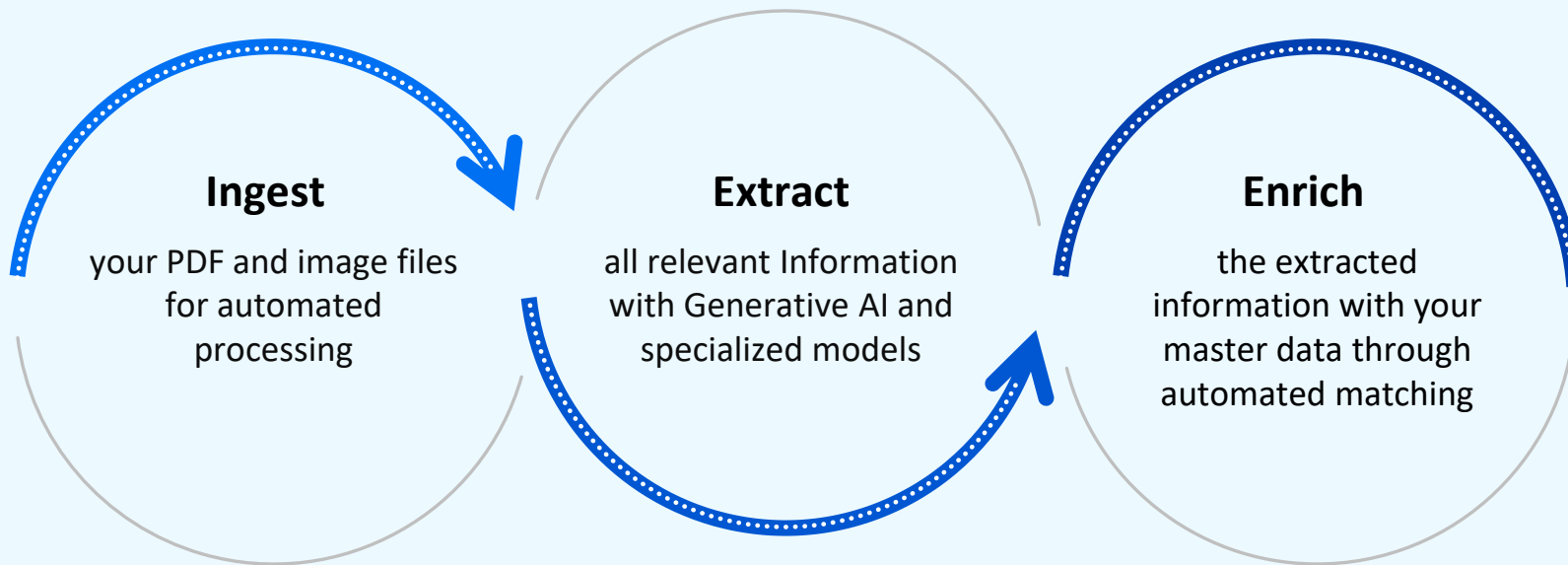
---

## **Topic 2: Document Information Extraction Premium Edition overview**



# Document Information Extraction

Intelligent document processing with SAP  
available on SAP Business Technology Platform



“

At Aspen Pumps, the Document Information Extraction service from SAP helps us to achieve ~90% of extraction accuracy for supplier invoices out of the box, laying a great foundation for much faster execution of Accounts Payable processes. It helps to save up to 350 hours monthly in invoice processing, so approximately £5000 per month.

- Jon Curtis, Group Business Systems Consultant, Aspen Pumps Group

 HAMBURG

 HomeMade

 REHAU

 AVAYA

 Walgreens Boots Alliance

 SAP  
Innovation  
Awards 2023

 SAP  
Innovation  
Awards 2022



# Document Information Extraction



## BASE

Automate core business document processing use cases in selected languages.

Use cases (supported languages)

- Invoice (15)
- Payment Advice (2)
- Purchase Order (3)
- Business Card (14)
- Templates (41)

## PREMIUM

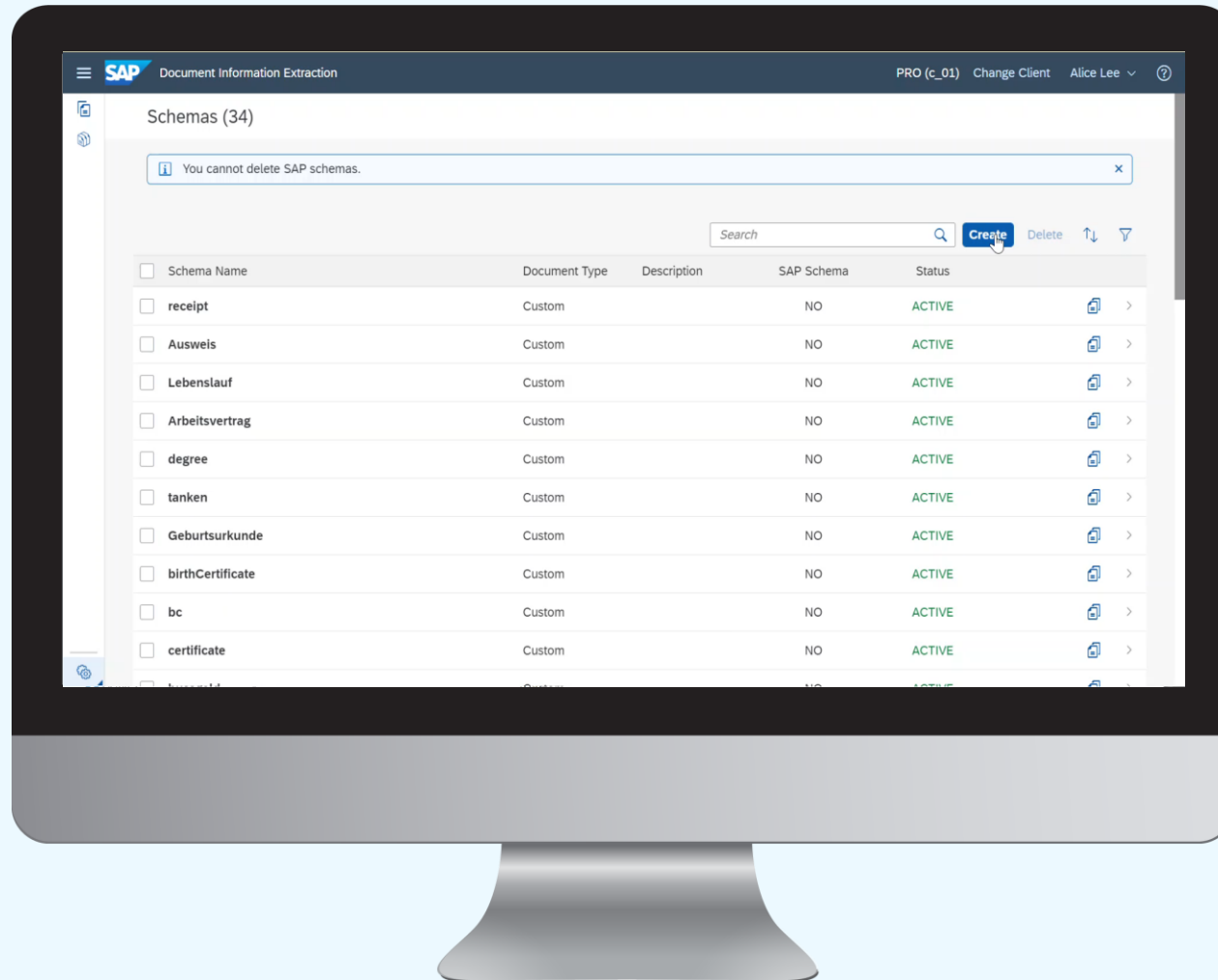
Automate all business document use cases in most languages end-to-end.

Out of box support for **40+ languages with easy extensions** for any kind of document.

- Auto extraction
- Language coverage
- Extensions of SAP Standard Schemas
- Outlook: Immediate Improvement

# Extract information from any document type with Document Information Extraction premium edition

Q4 2023



## Extract

Relevant information from any document type

## Customize

The list of fields that you want to extract based on your business needs

## Accelerate

No need to train custom model on own data and documents which might have various formats

# Scope



## Auto extraction

Onboarding / implementing new use cases

### Solution Description

- Provide access to Large Language Models (LLM) which are capable to process any kind of a document
- Use *Field Name* and *Description* to influence the prompt
- While creating a schema for a document processing use case configure the set-up type as *auto* (as opposed to *manual*)
- *Auto* enables to extract requested fields (entities) without the need to annotate a document

### Key Benefits

- Easy onboarding emerging use cases
- High flexibility with regard to the fields to be extracted
- No effort to annotate templates
- Short time-to-value



## Language coverage

Process any business document in more than 40 languages\*

### Solution Description

- Large Language Models (LLM) can process content in almost any kind of language
- This allows DOX to take full advantage of the language capabilities of in-house Business OCR service
- Combining Business OCR and LLMs allows now to process any business document type in more than 40 languages

- Global coverage
- Faster go-to-market

\* Actual language support will be subject to tests by the data science team and availability of test data

---

## Topic 3: Demo





---

## Topic 4: Roadmap



# Value-add capabilities and business benefits

## 2024 Roadmap

Outlook 2024



### Extensions of SAP Standard Schemas

Define any custom entity required for the specifics of a particular business process

#### Key Benefits

- Quick extensibility
- Higher business value



### Immediate Improvement

Automated learning based on user feedback. Feature will also enable “few shot” learning, i.e. providing vivid examples

- Better accuracy
- Higher automation rates



### Schema description

Add additional context to the schema of your document type; add additional instructions to guide the model in its extraction process (e.g., instructing the model to only extract something when special keywords appear)

- Better extraction results for complex cases (e.g. mixed documents)
- Higher accuracy



### Enhanced field description

Extend field description by 2x to provide a longer and more precise context for each field to be extracted

- Higher flexibility
- Influence extraction results



### Field grouping / multiple table support

Wrap fields into meaningful groups and extract information from complex table structures

- Categorize information
- Extract from nested tables
- Extract from multiple tables

# Wrap Up

---

~80 percent of data is unstructured. Great deal of business data is locked in business documents

SAP is at the forefront of adopting Generative AI in reliable, relevant and responsible way

SAP provides state-of-the-art GenAI capabilities to process unstructured documents of any kind.

## Enablement



### Assets:

Available on SAP BTP Trial - get started with [tutorials](#)!

State tuned with [SAP Community](#) for the GA!