



# Hands-on BTP AI

## Document Information Extraction Data Attribute Recommendation

**Name:** Tomaz Sauter / Joel Rodriguez

**Date:** April 2025

# Agenda

- Recap DOX + DAR | 20min
- Hands-on Document Info Extraction + Demo call APIs | 60min
- Hands-on Data Attribute Recommendation | 60min
- Next Steps | 10min

# SAP Business AI approach

Relevant

Reliable

Responsible

## Joule

A copilot that truly understands your business

### Embedded AI capabilities

SAP Cloud ERP

SAP Supply Chain  
Management

Human Capital  
Management

Spend Management  
and SAP Business  
Network

SAP Customer  
Relationship  
Management

SAP Business  
Technology Platform

**Your customized  
AI application**

## AI Foundation

on SAP Business Technology Platform

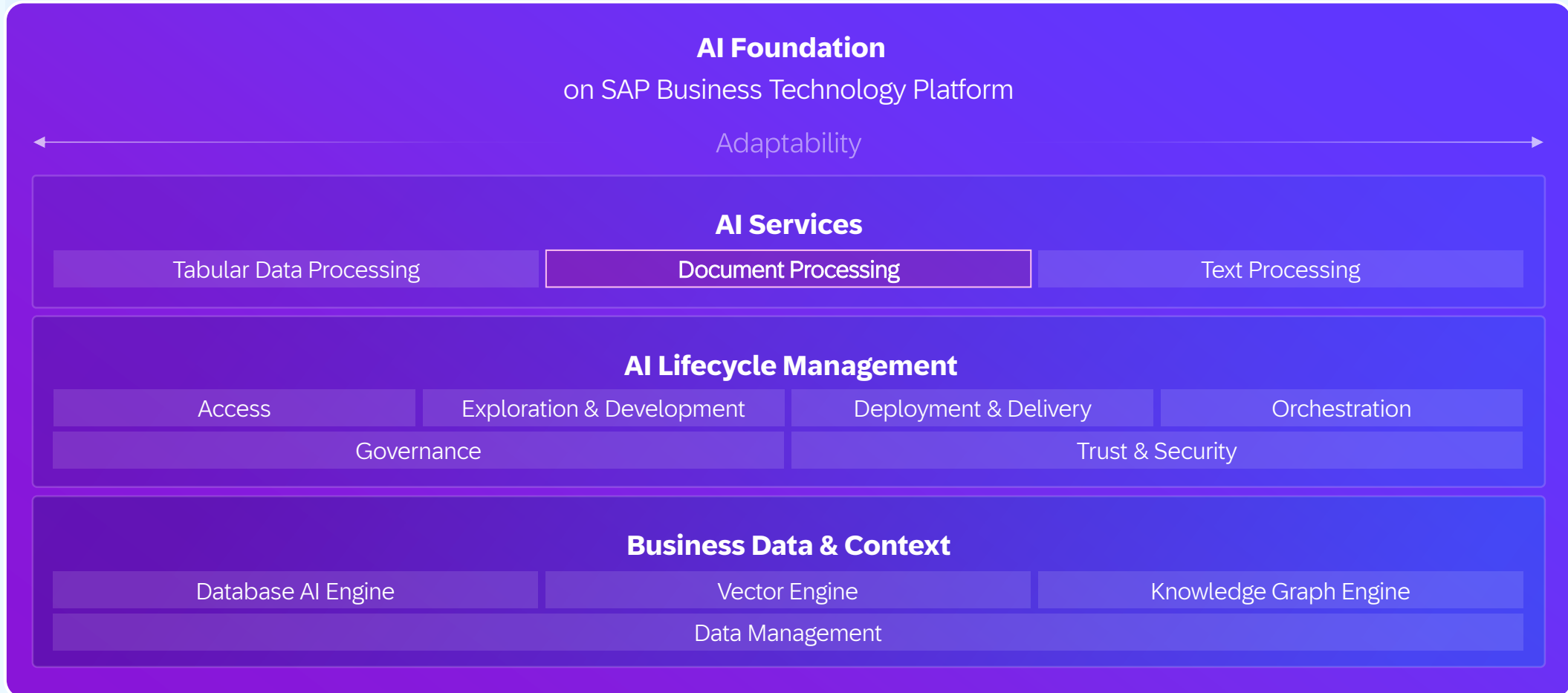
### AI ecosystem partnerships and investments



ANTHROPIC



# AI Foundation is a complete set of services for AI developers on SAP BTP



## 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*



# Problem: Large Amounts of Incoming Business Documents



**John,**  
AP-Clerk

**Time2Go Company**

**Your Company**  
M. Max Apfel-Birne  
Birnenstr. 24  
54321 Birnenstadt

**Contact:**  
Time2Go Company  
Am Apfelbau 13  
12345 Apfelstadt  
**Tel.** +1134/43215  
**Date:** 01.11.2013  
**Invoice n°:** 123456  
**Customer n°:** 0987

**Invoice**

Amount	Unit	Object	Price
1400	Pc.	Apples	\$1,684.-
120	Kg	Nuts	\$2.000
<b>Total Amount</b>			<b>\$ 3,684.-</b>
19% taxes incl.			

To pay until 11.12.2013

Time2Go Company  
Am Apfelbau 13  
12345 Apfelstadt

DE12 123412341234  
Apfelbank  
XYZ1234ZXY

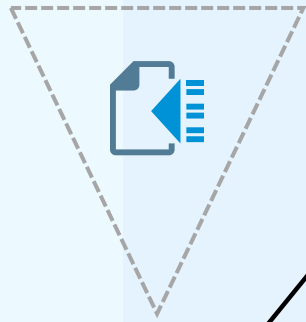
Company	Time2Go Company
Supplier ID	009753
Invoice N°	123456
Amount	3684.00
Currency	USD
Date	2013-11-01
Position 1	Amount: 1400 Unit: PC Amount: 1684.00 Description: Apples
Position 2	Amount: 120 Unit: KG Amount: 2000.00 Description: Nuts

Enrichment happens manually



**John,**  
spends much time  
on searching and  
matching the right  
information

# Solution: Document Information Extraction



John,  
AP-Clerk

**Time2Go  
Company**

**Your Company**  
M. Max Apfel-Birne  
Birnenstr. 24  
54321 Birnenstadt

**Contact:**  
Time2Go Company  
Am Apfelbau 13  
12345 Apfelstadt  
**Tel.** +1134/43215  
**Date:** 01.11.2013  
**Invoice n°:** 123456  
**Customer n°:** 0987

**Invoice**

Amount	Unit	Object	Price
1400	Pc.	Apples	\$1,684.-
120	Kg	Nuts	\$2.000
<b>Total Amount</b>			<b>\$ 3,684.-</b>

19% taxes incl.

To pay until 11.12.2013

Time2Go Company  
Am Apfelbau 13  
12345 Apfelstadt

DE12 123412341234  
Apfelbank  
XYZ1234ZXY

<b>Company</b>	Time2Go Company
<b>Supplier ID</b>	009753
<b>Invoice N°</b>	123456
<b>Amount</b>	3684.00
<b>Currency</b>	USD
<b>Date</b>	2013-11-01
<b>Position 1</b>	Amount: 1400 Unit: PC Amount: 1684.00 Description: Apples
<b>Position 2</b>	Amount: 120 Unit: KG Amount: 2000.00 Description: Nuts

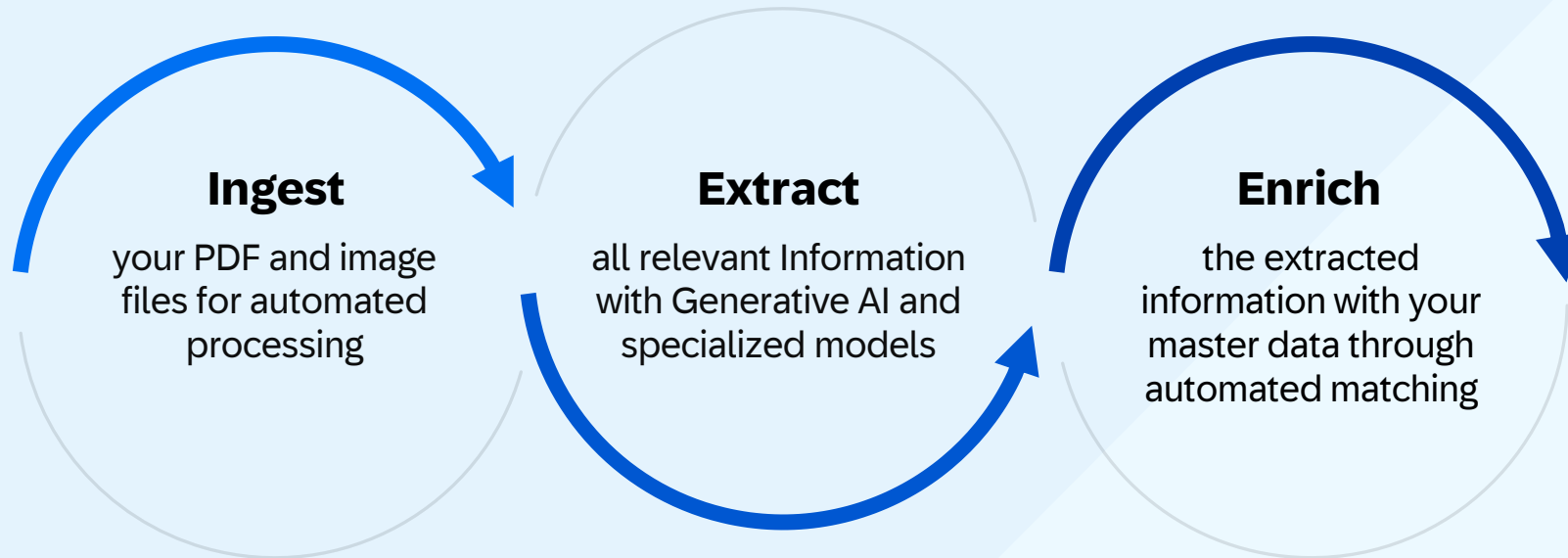
Enrichment  
happens  
automatically



**John,**  
Can now do more  
value creative tasks

# Document Information Extraction

Intelligent document processing with SAP available on SAP Business Technology Platform





# Benefit from Premium Capabilities

## Base AI

Automate core business document processing use cases in selected languages.



Supplier Invoice

Business Card

Payment Advice

Templates

Purchase Order

NEW

## Premium AI

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.



Base edition scope

Auto extraction based on Generative AI

Document Classification

Instant Learning Framework

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

## Extract

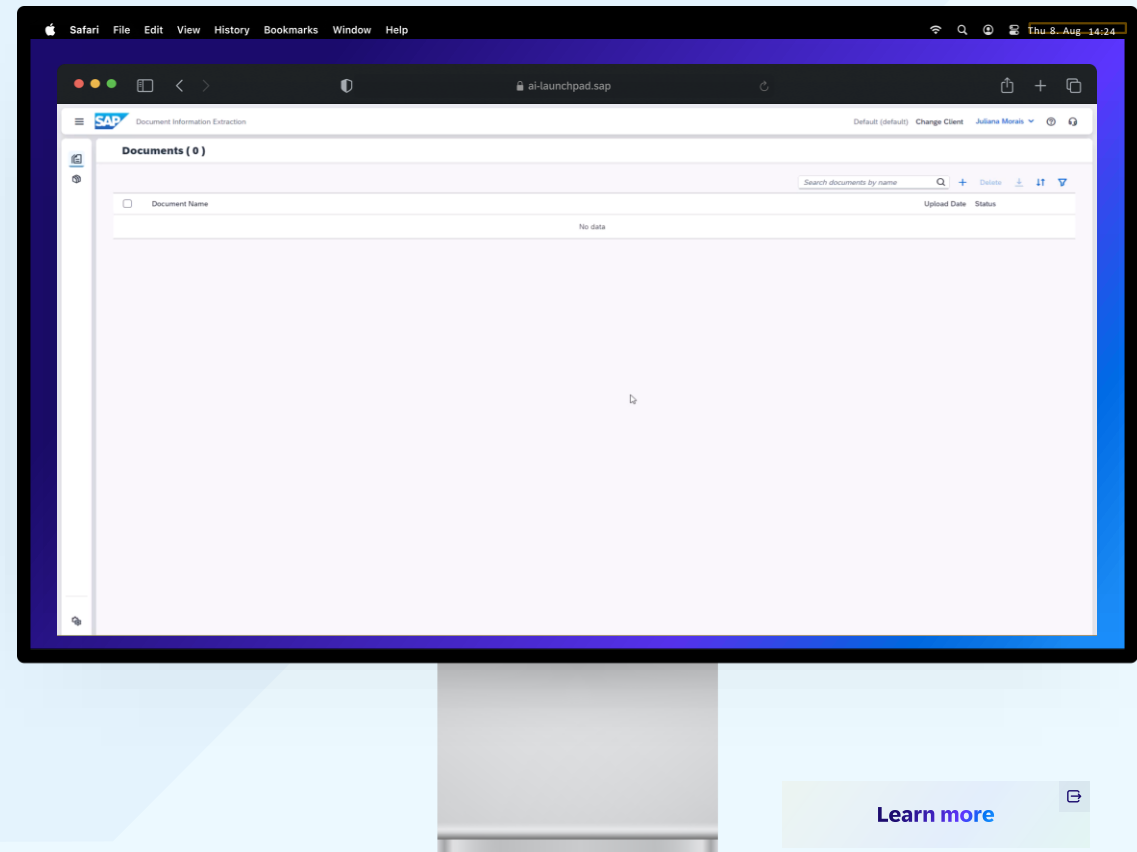
relevant information from any document type in 40+ languages with the help of Artificial Intelligence

## Customize

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

## Accelerate

your processes without training custom ML model or managing thousands of templates



[Learn more](#)



SAP Build

InvoiceProcess-Simple - SAP Build

Process Automation Trigger Form

Invoice1.pdf

Invoice2.pdf

build-t1kx1uh1.sap-process-automation.cfapps.eu10.hana.ondemand.com/studio/#9a78f565-ee2b-4725-803a-628c6ef0db13

Share

Star

Help

Settings

Profile

SAP

Build Process Automation

InvoiceProcess-Simple

Editable

Overview

InvoiceAutomation

Invoice

Invoice

ApprovalForm

InvoiceTrigger

Release

Save

InvoiceTrigger

Load Invoice Data

Vendor \*  
Enter Text T

Contact Mail \*  
Enter Text T

Total Ammount  
Enter a Value #

Register Day

InvoiceTrigger

Description  
Enter Form Description  
0/1024

Design Console (0)

Test Console (0)

Variables (0)

Unattended mode

Taskbar

System Tray

# 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

[Learn more](#)



 HAMBURG  HomeMade

 REHAU

AVAYA

 SAP  
Innovation  
Awards 2022

 SAP  
Innovation  
Awards 2023

 Walgreens Boots Alliance

## Document Information Extraction

Intelligent document processing with SAP available on SAP Business Technology Platform

*“The real ‘wow’ moment happened when we were able to see and appreciate the KPIs related to the accuracy of the Document Information Extraction service. We expected to achieve around 70% to 80% automatic loading from PDF invoices. Then, we realized the new solution was actually delivering 90% to 95% text capture from invoices. With that, we achieved a savings of more than 680 hours of manual work each month.”*

**Davide Ardizio**

CIO, De Agostini Publishing S.p.A.

**54,000**

paper-based vendor invoices  
processed in a year

**>91%**

Of PO-referred invoices  
automatically processed

**>680**

hours saved per month

[Learn more](#)





## Customer Reference

Industry:  
Public Sector

### Featured Solutions and Services:

SAP Business Technology Platform  
Document Information Extraction, premium edition  
Business Entity Recognition  
Generative AI hub in SAP AI Core  
SAP Analytics Cloud  
SAP Business Integrity Screening  
SAP S/4HANA

*“ With the platform designed to distribute aid to artists out of the special fund for cultural events, we could help support the quick restart of cultural events throughout Germany, which is so important for our communities and social cohesion. “*

### Carsten Brosda

Minister of Culture and Media, Free  
and Hanseatic City of Hamburg

**6 million**

documents automatically  
evaluated and classified

**99,000**

cultural aid applications  
filed on the platform so far

**2 weeks**

to deliver time-critical aid application platform

[Learn more](#)



# **AI Business Services – Data Attribute Recommendation**



# Business Challenge - Matching Information done Manually

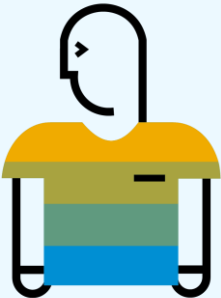


**Bob,**  
Master Data  
Team

Material Table

Description	Manufacturer	Price	Category	Subcategory
1875 W Conair Hair Dryer	Conair	19.99	Health, Fitness & Beauty	Personal Care & Beauty
40x to 1000x - LED Illumination - Monocular Head	Celestron	369.99	Cameras & Camcorders	Binoculars, Telescopes & Optics
Amazons 6" Paperwhite display	Amazon	289.99	Computers & Tablets	E-Readers & Accessories
Yellow - Laser - 1500 Page - 1 Each	Brother	50.99	Connected Home & Housewares	Office & School Supplies
Take the fight to the next level	Activision	19.99	Video Games	Xbox 360
Compatible with select electronic devices	alkaline chemistry	7.99	Connected Home & Housewares	Housewares
From our expanded online assortment; compatible with most 1989-2000 Ford, Lincoln and Mercury vehicles	Metra	16.99	Car Electronics & GPS	Car Installation Parts & Accessories

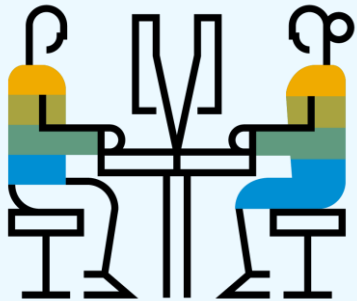
Manual  
matching  
and  
insertion of  
information



**Bob,**  
spends much time  
on repetitive  
tasks



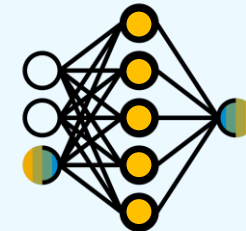
# AI Approach - Train a model with historical / labelled data



Material Table

Description	Manufacturer	Price	Category	Subcategory
1875 W Conair Hair Dryer	Conair	19.99	Health, Fitness & Beauty	Personal Care & Beauty
40x to 1000x - LED Illumination - Monocular Head	Celestron	369.99	Cameras & Camcorders	Binoculars, Telescopes & Optics
Amazons 6" Paperwhite display	Amazon	289.99	Computers & Tablets	E-Readers & Accessories
Yellow - Laser - 1500 Page - 1 Each	Brother	50.99	Connected Home & Housewares	Office & School Supplies
Take the fight to the next level	Activision	19.99	Video Games	Xbox 360
Compatible with select electronic devices	alkaline chemistry	7.99	Connected Home & Housewares	Housewares
From our expanded online assortment; compatible with most 1989-2000 Ford, Lincoln and Mercury vehicles	Metra	16.99	Car Electronics & GPS	Car Installation Parts & Accessories

Training data

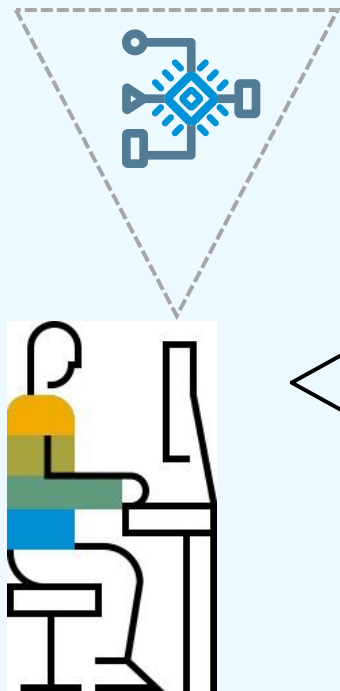


Data Attribute Recommendation

\*model trained\*



# Solution – Automation with Data Attribute Recommendation

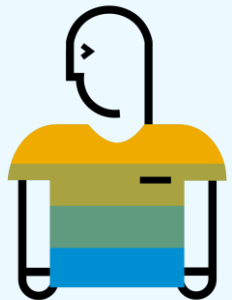


**Bob,**  
Master Data  
Team

Material Table

Description	Manufacturer	Price	Category	Subcategory
1875 W Conair Hair Dryer	Conair	19.99	Health, Fitness & Beauty	Personal Care & Beauty
40x to 1000x - LED Illumination - Monocular Head	Celestron	369.99	Cameras & Camcorders	Binoculars, Telescopes & Optics
Amazons 6" Paperwhite display	Amazon	289.99	Computers & Tablets	E-Readers & Accessories
Yellow - Laser - 1500 Page - 1 Each	Brother	50.99	Connected Home & Housewares	Office & School Supplies
Take the fight to the next level	Activision	19.99	Video Games	Xbox 360
Compatible with select electronic devices	alkaline chemistry	7.99	Connected Home & Housewares	Housewares
From our expanded online assortment, compatible with most 1989-2000 Ford Lincoln and Mercury vehicles	Metra	16.99	Car Electronics & GPS	Car Installation Parts & Accessories

Predictions  
for  
customer  
specific  
fields

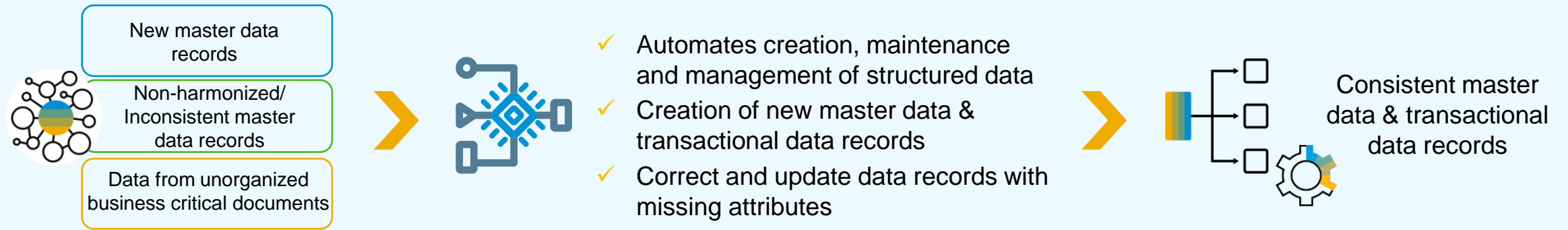


**Bob,**  
Can now do more  
value creative  
tasks



# What is Data Attribute Recommendation ?

Automate master data / transactional data management tasks



**Data Attribute Recommendation helps to classify entities such as products, stores, sales orders and users into multiple classes, using free text, numbers and categories as input.**



Automate and speed up master data creation and maintenance

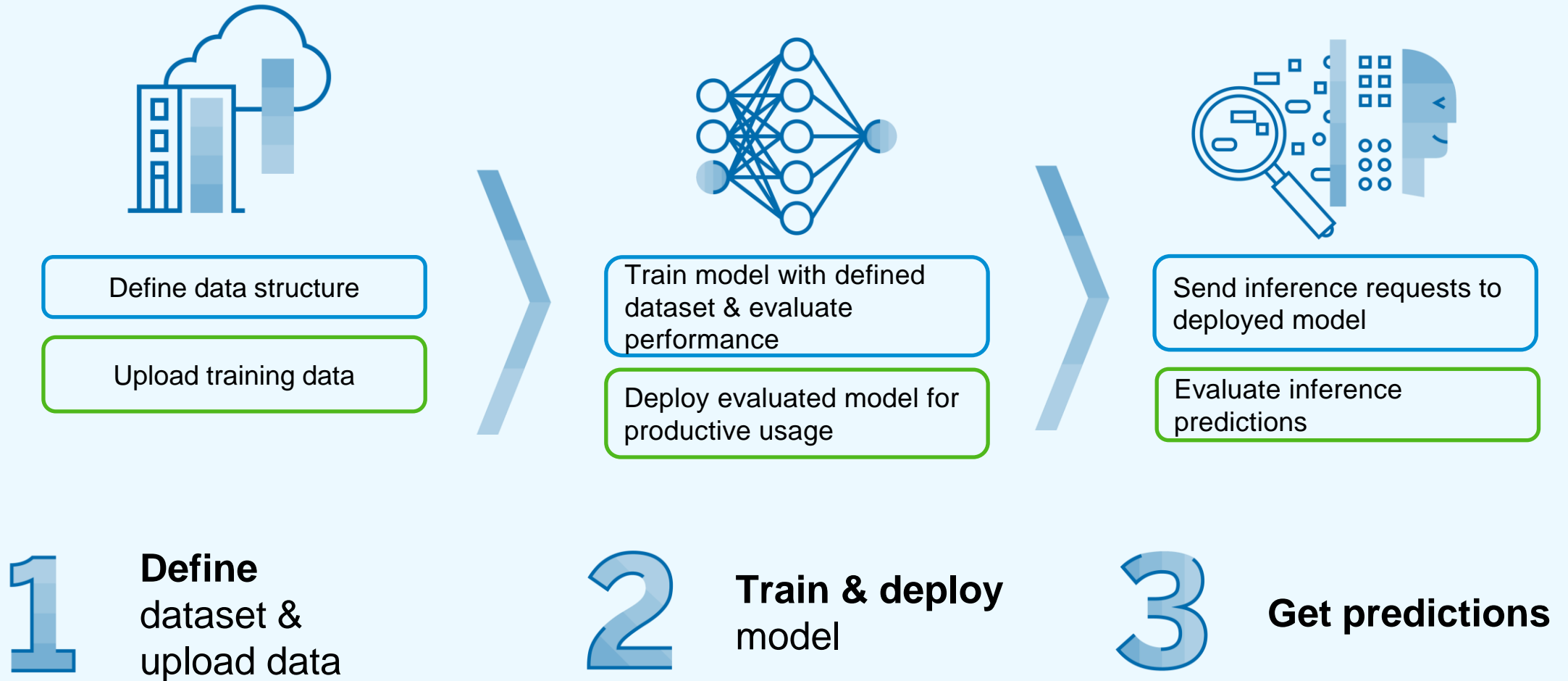


Gain easier and faster master data insights



Reduce errors and manual efforts when maintaining data records

# How Data Attribute Recommendation works



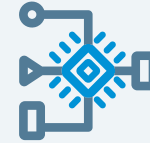
# Automatic Prediction of missing Master Data fields for New Product Introduction



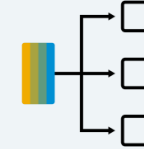
New material creation request  
(CSV, Web based)



Minimum number of  
information input by  
the business user



Transfer input data  
either online or batch  
mode



Predict fields that can  
be integrated to  
material master

## Business Challenge



- Constant **new product release** to the market
- **Lack of holistic view** of the master data needed by **different business functions**
- **Inconsistent representation** of master data across the entire organization



## Solution



### Data Attribute Recommendation

- Provides master data field prediction to the business user
- **Retrainable Machine learning** models that can be retrained at frequent intervals
- Can be used for business partner maintenance, work center and asset maintenance, custom field **value prediction**



## Benefits, Value



- ✓ **Reducing** the manual maintenance and verification
- ✓ **One voice of master data** across the organization
- ✓ **Improve quality of master data** to facilitate more efficient planning and execution

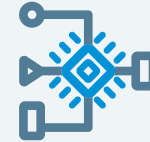
# Intelligent Sales Order Creation Processing



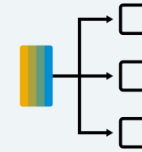
Inbound Sales Orders  
(Online / EDI / Email)



Incomplete Sales  
orders created by  
service agent / Bot



Provide incomplete  
fields as input



Predict missing  
and complete sales  
order creation  
process

## Business Challenge



- **Incomplete and overdue sales order** due to missing information
- **High manual efforts** needed to retrieve the missing information
- **Delay in order fulfillment** in the order to delivery process



## Solution



### Data Attribute Recommendation

- ✓ **Incomplete fields** to be fed as input to Data Attribute Recommendation service
- ✓ Can be combined **with rule based procedures**
- ✓ **Exception classification** of sales orders **for automated processing**



## Benefits, Value



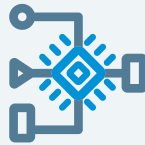
- ✓ **Significantly reduce** manual efforts for order completion
- ✓ **Expedite** overall sales order processing
- ✓ **Leverage** extensive cost saving potential

\*The same solution can be used in the case of purchase order creation

# Commodity Code Prediction



New Commodity,  
no Commodity Code



Data Attribute  
Recommendation



Commodity  
Code



New Commodity is  
Imported / Exported

## Business Challenge



- **International trade demands commodity codes**
- **No correct commodity code**
  - **Rejection** of assignment goods by customs
  - **Waiting on borders** = **high costs** for the organization
- **Maintaining** such master data
  - is supported by rules
  - leads to **inconsistencies & left-overs** for manual processing



## Solution



- Data Attribute Recommendation** can predict:
- ✓ **Correct values by learning the logic behind** complex fields (*such as the commodity code*)
  - ✓ **Dependencies** to multiple data attributes



## Benefits, Value

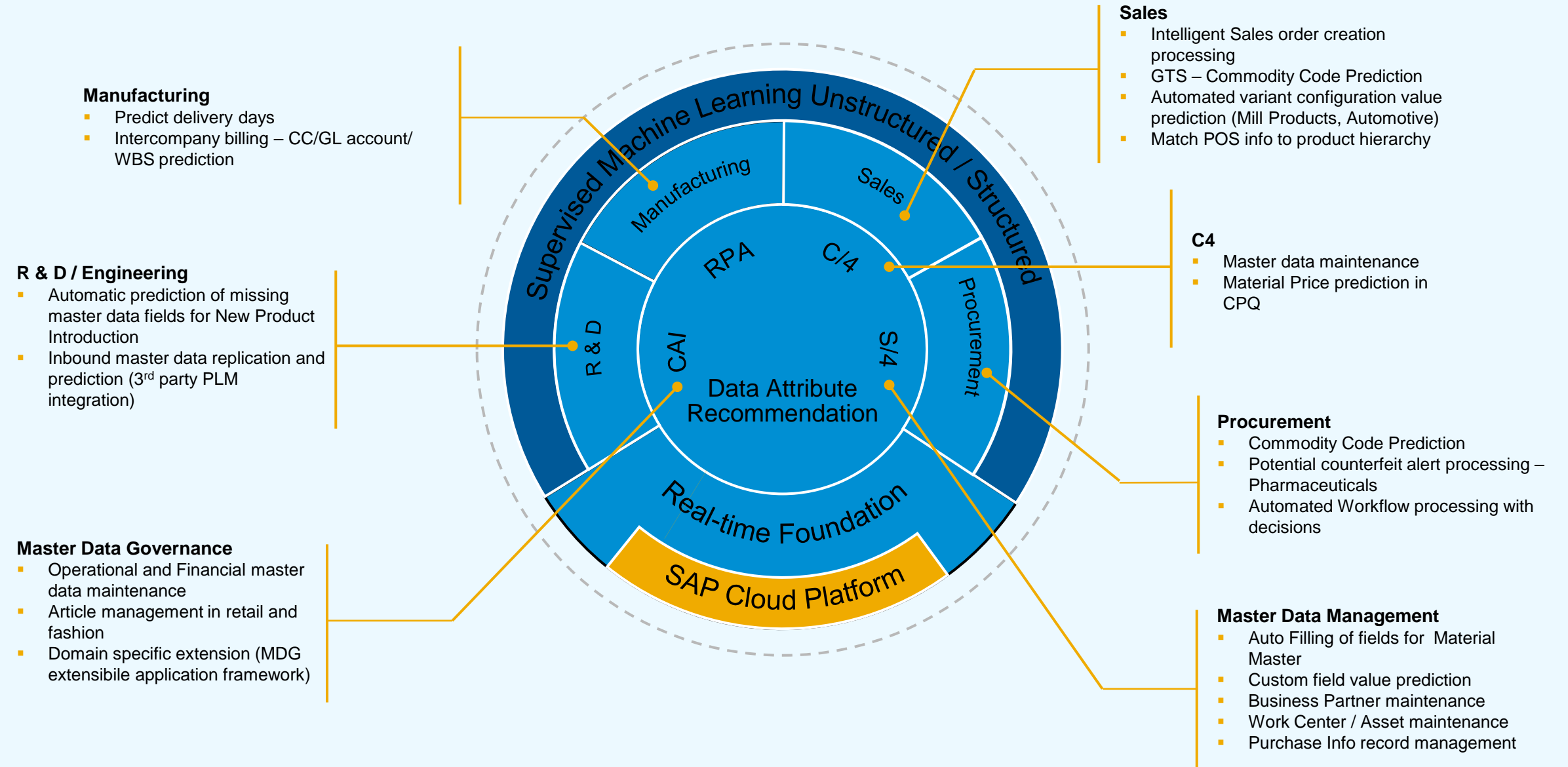


- ✓ **Reduces manual efforts** in commodity code assignment
- ✓ **Increases the validity** of assigned commodity codes
- ✓ **Decreased costs:** caused by labor & mal-assignment of commodity codes



# Intelligent Master and Transaction Data Enrichment

## Further Potential Use Cases



SAP AI Business Services

sapai-lwbk60xn4a.dispatcher.eu3.hana.ondemand.com/index.html#/dar

<

SAP

Data Attribute Recommendation

Global Platform & Technologies Presales

## Demo Description

Data Attribute Recommendation is an SAP Cloud Platform service that applies machine learning to match and classify data records, making more efficient and higher-quality master data management a reality. Data Attribute Recommendation can be used, for example, to assign incoming products to an existing product hierarchy. Another case could be an enrichment of your data, by filling the missing attributes of a data record.

For this demo, we have used a sample dataset which could be found in the sample input section. To run the demo, pick any row data (price, manufacturer, description) and enter into the input section to get the existing product hierarchy in F1, F2 and F3 Readables with confidence values for the submitted data.

<

Input

Enter Description

Enter Manufacturer

Enter Price

Submit

>

Sample Input

Name	Category	Confidence
No data		

© 2024

30

# DAR Demo – Sample Data, UI

Input

Compatible with select electronic devices; AAA size; DURALOCK Power Preserve technology; 8-pack

Duracell

7.49

Submit

Sample Input

Name	Category	Confidence
F1_readable	Connected Home & Housewares	99.97%
F1_readable	Appliances	0.01%
F2_readable	Housewares	99.69%
F2_readable	Household Essentials	0.11%
F3_readable	Household Batteries	98.4%
F3_readable	Camera Batteries & Power	0.53%

description	manufacturer	price	F1_readable	F2_readable	F3_readabl e
Compatible with select electronic devices; AAA size; DURALOCK Power Preserve technology; 4-pack	Duracell	5.49	Connected Home & Housewares	Housewares	Household Batteries
Long-lasting energy; DURALOCK Power Preserve technology; for toys, clocks, radios, games, remotes, PDAs and more	Duracell	5.49	Connected Home & Housewares	Housewares	Household Batteries
Compatible with most vehicles; designed for the installation of an aftermarket radio	Metra	10.99	Car Electronics & GPS	Car Installation Parts & Accessories	Car Audio Installation Parts
Designed for use with most Par-56 can gel frames; red, blue, yellow and green gel filters	American DJ	12.99	Musical Instruments	DJ & Lighting Equipment	DJ Lighting & Effects
ALTEC 4000 mAh Power Bank: Compatible with most USB-enabled mobile devices; USB interface; built-in lithium-ion battery; up to 500 hours of standby time; overload protection; soft rubberized finish	Altec	29.99	Cell Phones	Cell Phone Accessories	Cell Phone Batteries & Power

# Improved master data quality through automation to increase accuracy and lead with Intelligent Processes

## Business Challenge

---

Ordering new spare parts for the first time, such as particular circuit breaker or drill, was often a significant effort for business users at Severstal, Russian steel and steel related mining company saw returns on approximately 21% of all new material requests due to the incorrectly chosen material classes

## Solution

---

Severstal-Infokom, the company's IT arm, sought a **novel approach to support business users** when they filled out a new request. It decided on Machine Learning to improve the process of entering new materials into the material master data in an intelligent and automated way

## Benefits

---

- ✓ **AI enabled** software landscape that uses machine learning to help **reduce business user errors**
- ✓ **Reduces** the effort for new material creation by **approximately 20%**
- ✓ **Automates** key processes in master data management as well as master data governance functions for **procured materials**



---

Industry : Mill Products  
Country : Russia

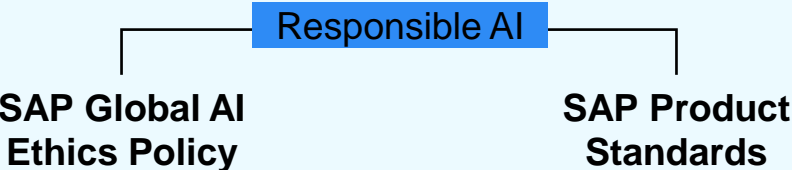


# Data privacy with generative AI

➤ SAP pursues enterprise-ready partner agreements for generative AI technology with multiple vendors that cater for data privacy and isolation.

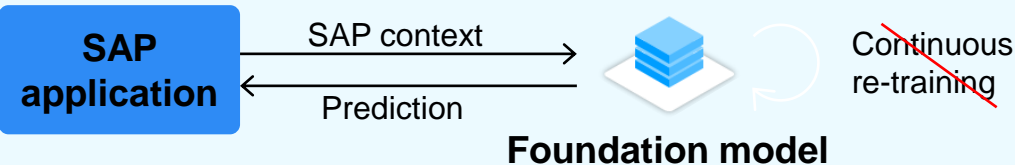
## RESPONSIBLE AI

SAP’s use of **generative AI** follows the same **principles of business AI** and the same rigorous development and responsible AI review processes.



AI Ethics at SAP

## DATA USAGE



For predictions, customer data and context is required to receive meaningful results.

However, SAP ensures that **no customer data is used** by third-party vendors to train foundation models.

Data Usage Guidelines



# Thank you

