CAMUNDA CON LIVE

Intelligent Process Automation with Camunda and Amazon Comprehend

Dr. Marigianna Skouradaki, Novatec Consulting GmbH

NOVATEC

Intelligent Process Automation with AWS and Camunda

Dr. Marigianna Skouradaki

23.09.2021

CAMUNDA | Platinum Partner CERTIFIED

Dr. Marigianna Skouradaki

Data Intelligence Consultant &

Certified Camunda Engineer

- BPM and data analysis
- Algorithms on graph similarities
- Benchmark for modern Workflow Engines
- Intelligent Process Automation

- Active in Camunda und Machine Learning Projects
- Trainer of the Data Science Training

Takeaways

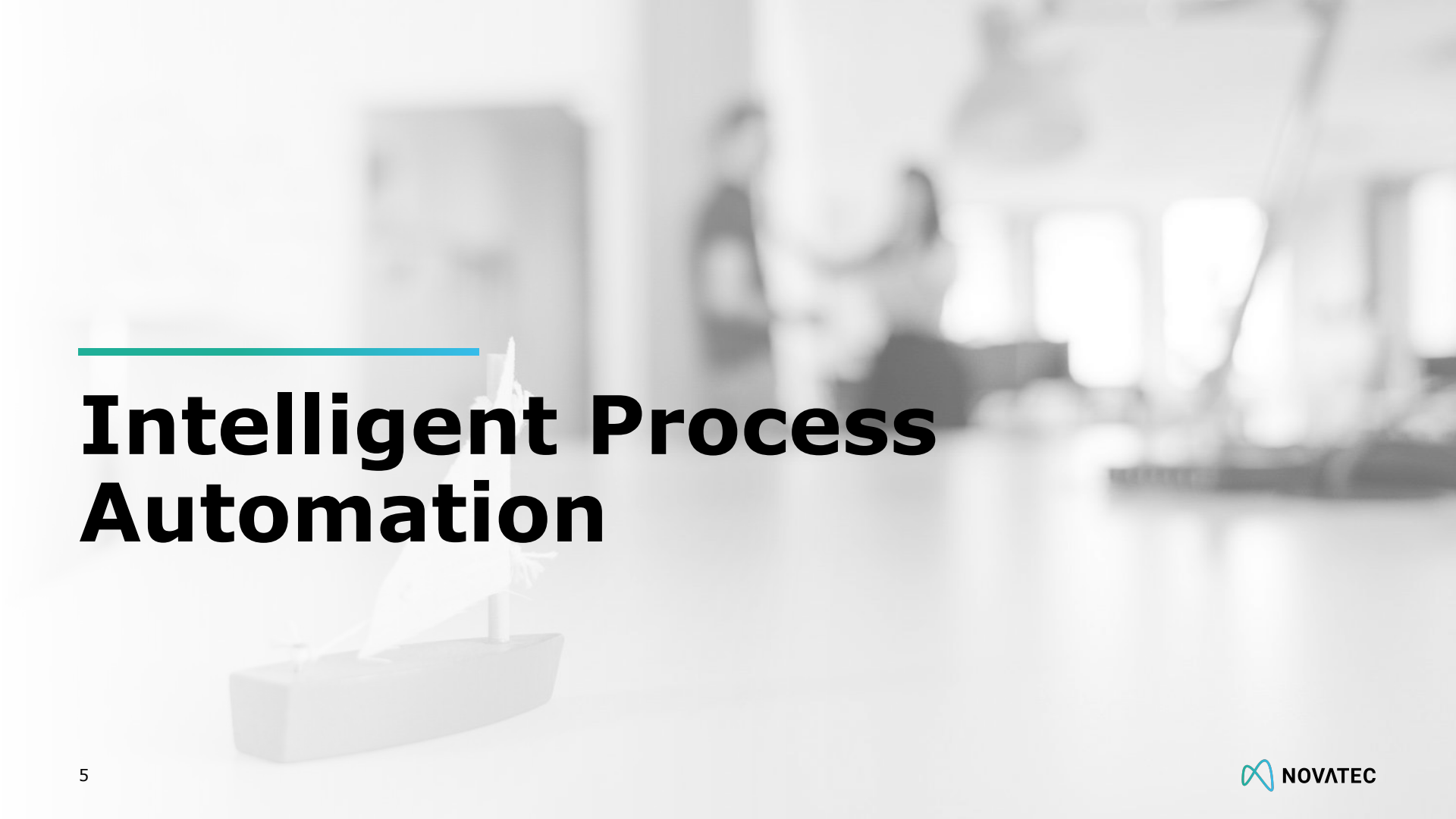
What is Intelligent Process Automation?

Why now?

How to build an intelligent process?

Agenda

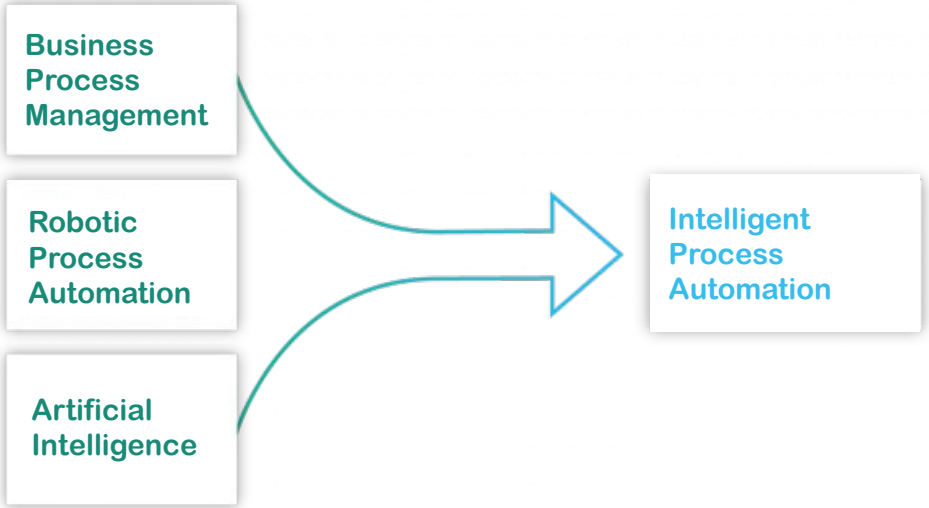
1. Intelligent Process Automation
2. Intelligent Costs Submission Process
3. Demo



Intelligent Process Automation

NOVATEC

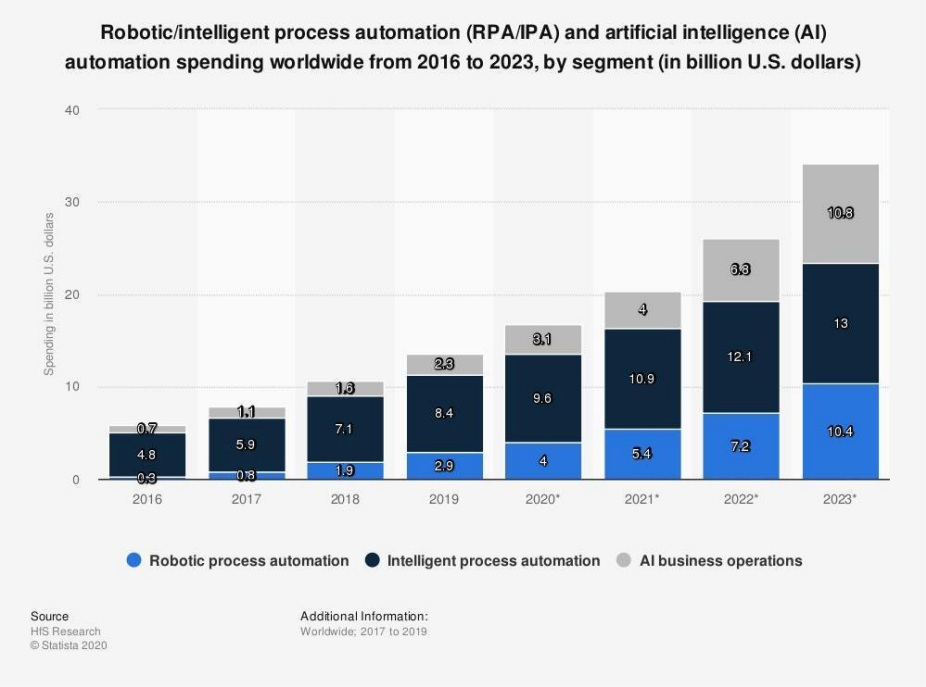
Intelligent Process Automation



https://www.novatec-gmbh.de/beratung/intelligente-prozessautomatisierung/

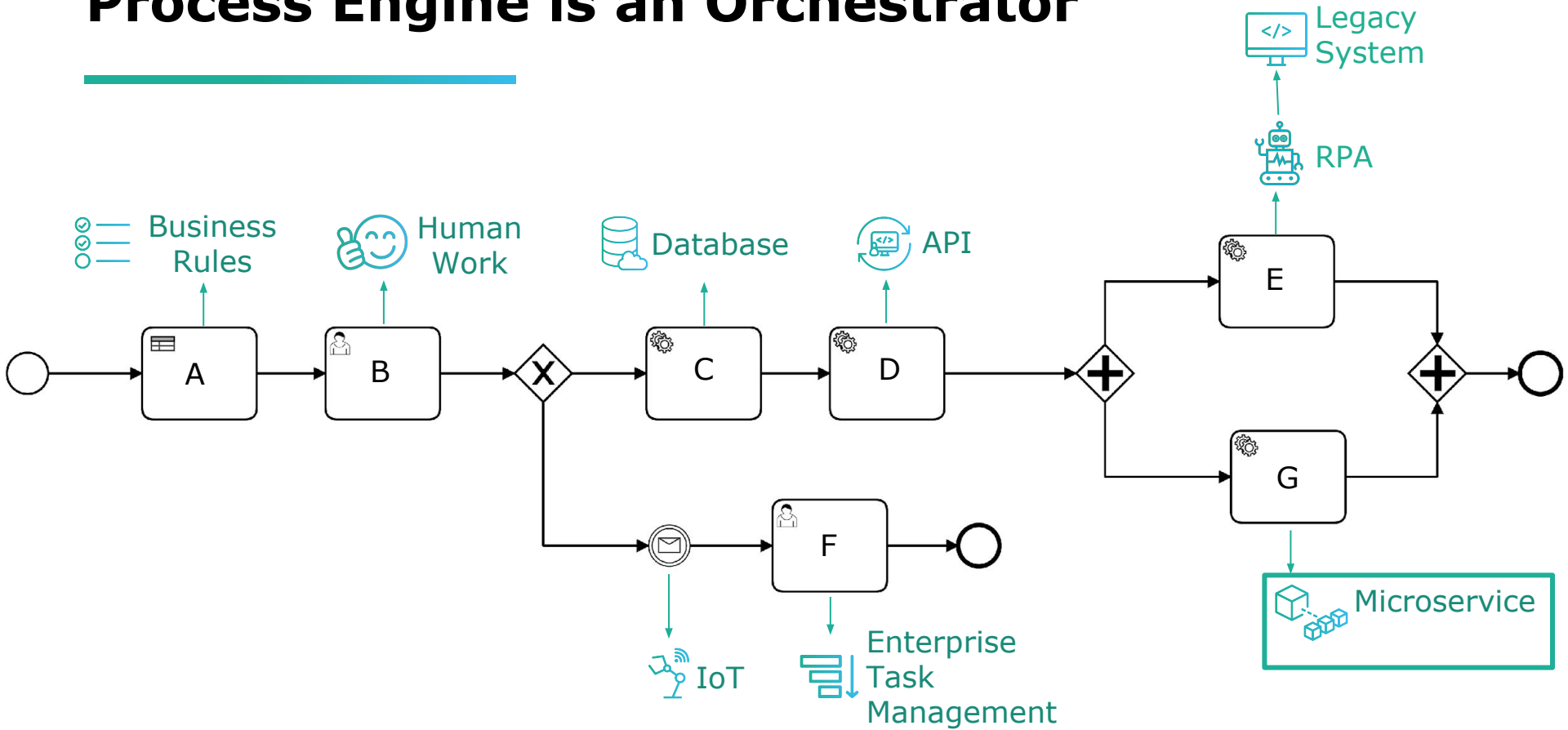
“The Future is Both Automated and Intelligent“ (Forbes)

- Camunda closed a $100 Million Series B Funding Round
- UiPath $10 Billion is the most valuable Software-Start-up of Europe
- SAP bought Signavio for 1B€
- IBM introduced Watson Orchestrate in 2021 and bought MyInvenio



NOVATEC

Process Engine is an Orchestrator



NOVATEC

Machine Learning Applications in BPM

Detection

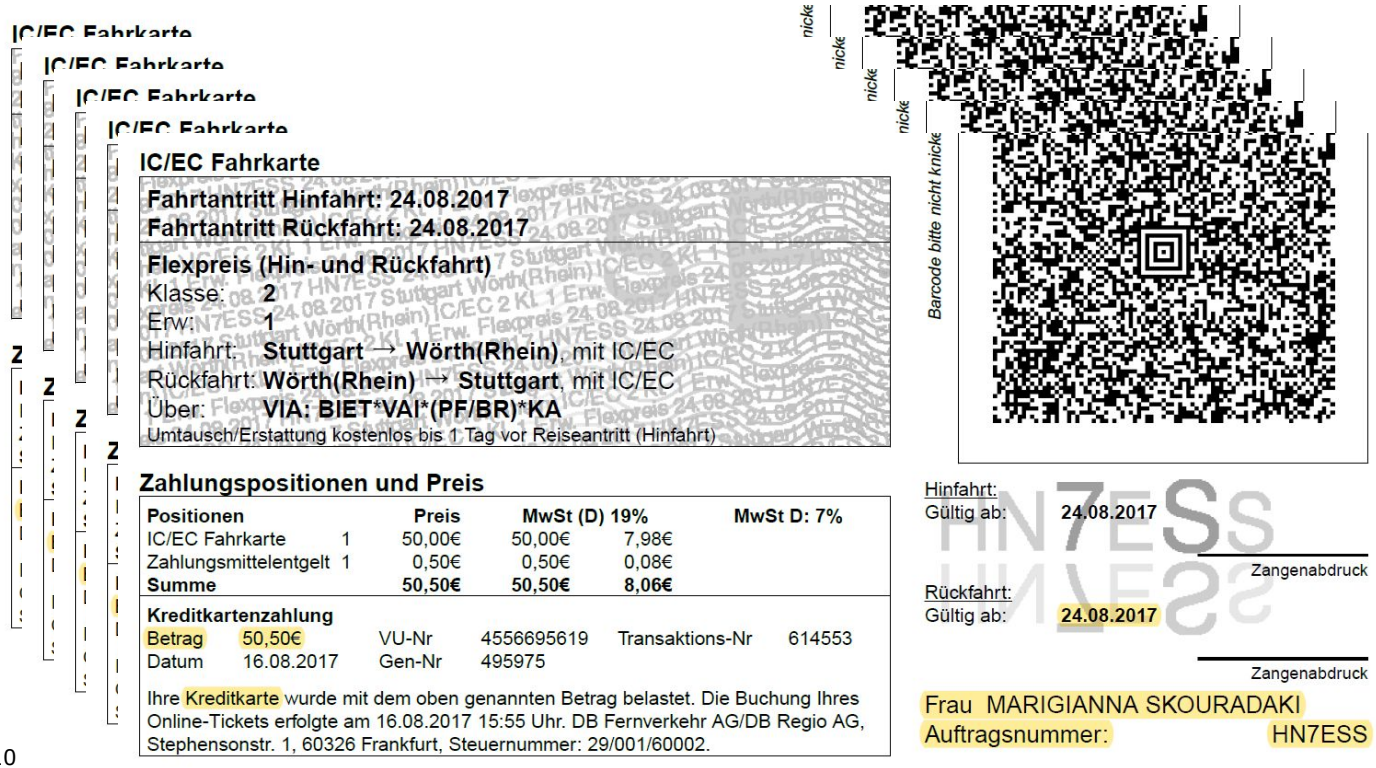
Analysis

Analysis

Systems

Detection

Example: Intelligent Document Analysis



IC/EC Fahrkarte

Fahrtantritt Hinfahrt: 24.08.2017
Fahrtantritt Rückfahrt: 24.08.2017

Flexpreis (Hin- und Rückfahrt)
Klasse: **2**
Erw: **1**
Hinfahrt: **Stuttgart → Wörth(Rhein)**, mit IC/EC
Rückfahrt: **Wörth(Rhein) → Stuttgart**, mit IC/EC
Über: **VIA: BIET*VAI*(PF/BR)*KA**
Umtausch/Erstattung kostenlos bis 1 Tag vor Reiseantritt (Hinfahrt)

Zahlungspositionen und Preis

Positionen		Preis	MwSt (D) 19%	MwSt D: 7%
IC/EC Fahrkarte	1	50,00€	50,00€	7,98€
Zahlungsmittelentgelt	1	0,50€	0,50€	0,08€
Summe		**50,50€**	**50,50€**	**8,06€**

Kreditkartenzahlung

Betrag	50,50€	VU-Nr	4556695619	Transaktions-Nr	614553
Datum	16.08.2017	Gen-Nr	495975		

Ihre Kreditkarte wurde mit dem oben genannten Betrag belastet. Die Buchung Ihres Online-Tickets erfolgte am 16.08.2017 15:55 Uhr. DB Fernverkehr AG/DB Regio AG, Stephensonstr. 1, 60326 Frankfurt, Steuernummer: 29/001/60002.

Barcode bitte nicht knicken

Hinfahrt:
Gültig ab: 24.08.2017

Zangenabdruck

Rückfahrt:
Gültig ab: 24.08.2017

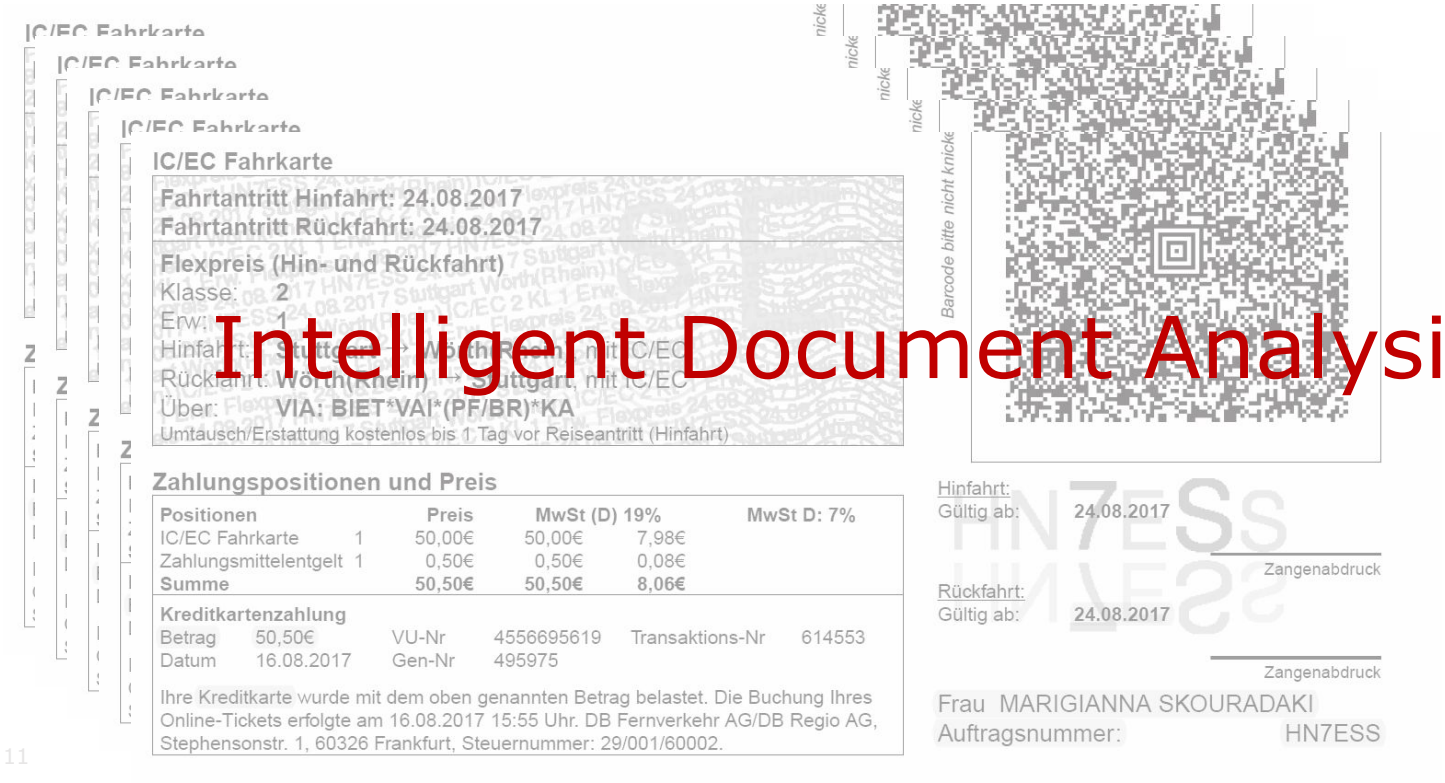
Zangenabdruck

Frau MARIGIANNA SKOURADAKI
Auftragsnummer: HN7ESS

- Invoices
- Receipts
- Handwritten files
- Contracts…

NOVATEC

Example: Intelligent Document Analysis

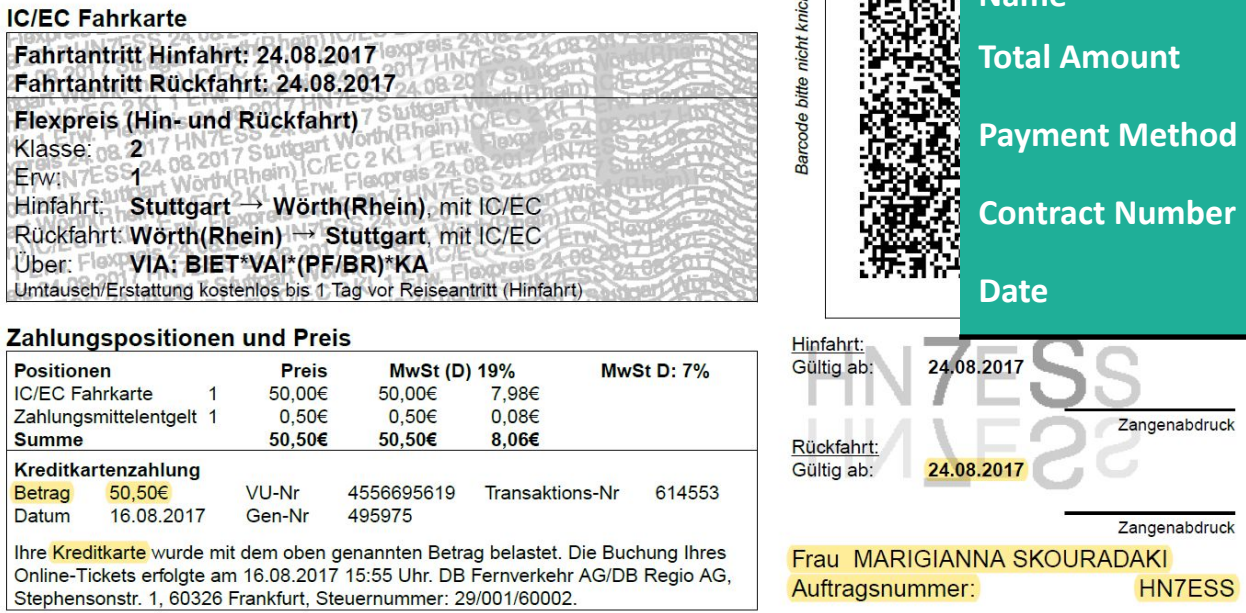


Intelligent Document Analysis

- Invoices
- Receipts
- Handwritten files
- Contracts…

NOVATEC

Example: Intelligent Document Analysis



IC/EC Fahrkarte

Fahrtantritt Hinfahrt: 24.08.2017
Fahrtantritt Rückfahrt: 24.08.2017

Flexpreis (Hin- und Rückfahrt)
Klasse: **2**
Erw: **1**
Hinfahrt: **Stuttgart → Wörth(Rhein)**, mit IC/EC
Rückfahrt: **Wörth(Rhein) → Stuttgart**, mit IC/EC
Über: **VIA: BIET*VAI*(PF/BR)*KA**
Umtausch/Erstattung kostenlos bis 1 Tag vor Reiseantritt (Hinfahrt)

Zahlungspositionen und Preis

Positionen		Preis	MwSt (D) 19%		MwSt D: 7%
IC/EC Fahrkarte	1	50,00€	50,00€	7,98€	
Zahlungsmittelentgelt	1	0,50€	0,50€	0,08€	
Summe		**50,50€**	**50,50€**	**8,06€**	

Kreditkartenzahlung

Betrag	50,50€	VU-Nr	4556695619	Transaktions-Nr	614553
Datum	16.08.2017	Gen-Nr	495975		

Ihre Kreditkarte wurde mit dem oben genannten Betrag belastet. Die Buchung Ihres Online-Tickets erfolgte am 16.08.2017 15:55 Uhr. DB Fernverkehr AG/DB Regio AG, Stephensonstr. 1, 60326 Frankfurt, Steuernummer: 29/001/60002.

Barcode bitte nicht knicke

Hinfahrt:
Gültig ab: 24.08.2017

Zangenabdruck

Rückfahrt:
Gültig ab: 24.08.2017

Zangenabdruck

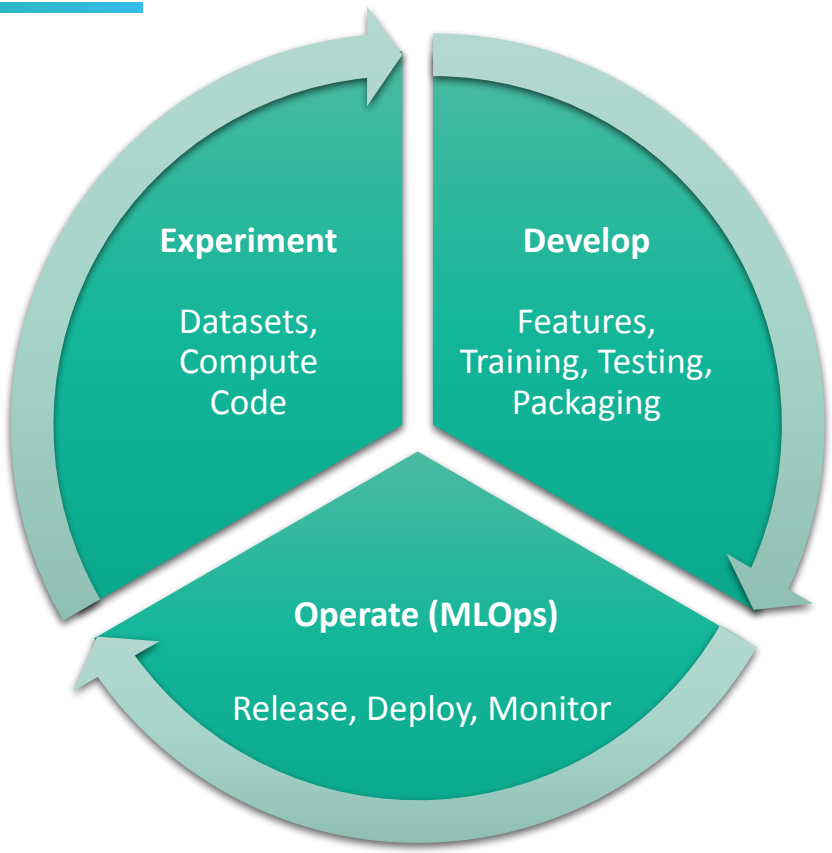
Frau MARIGIANNA SKOURADAKI
Auftragsnummer: HN7ESS

Name	MARIGIANNA SKOURADAKI
Total Amount	50,50€
Payment Method	CREDIT CARD
Contract Number	HN7ESS
Date	24.08.2017

Key-Value Pairs

NOVATEC

Machine Learning Lifecycle

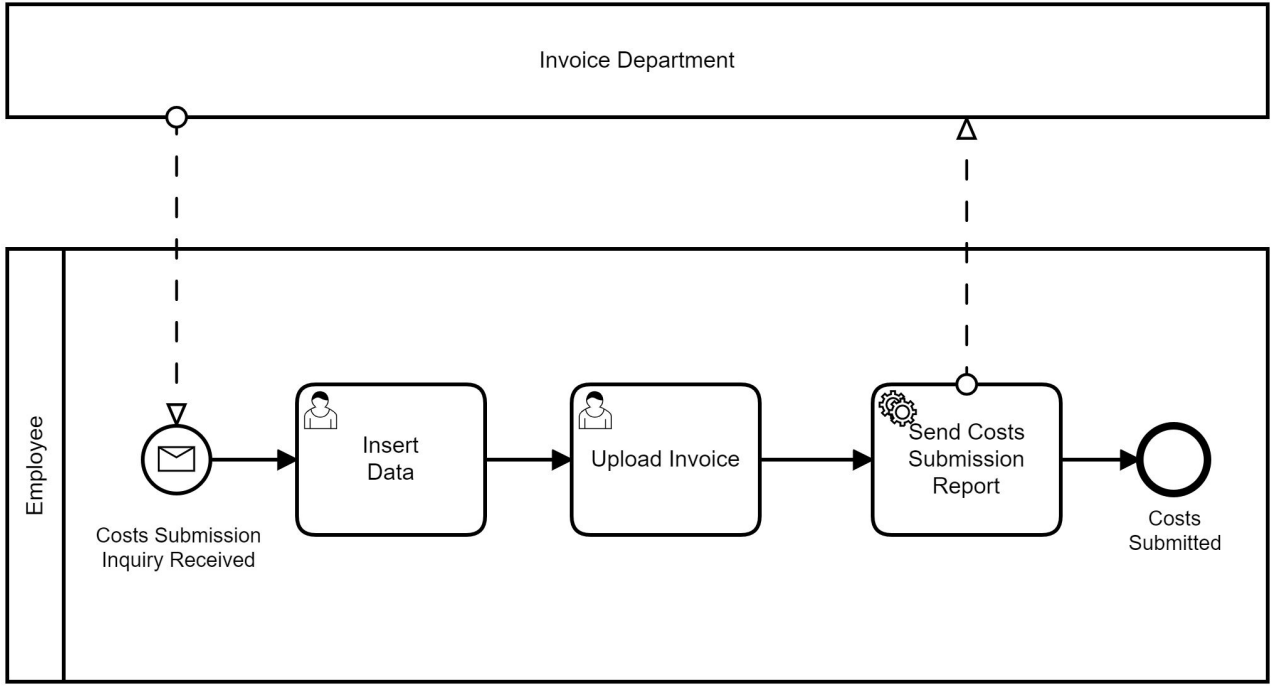


MLaaS

NOVATEC

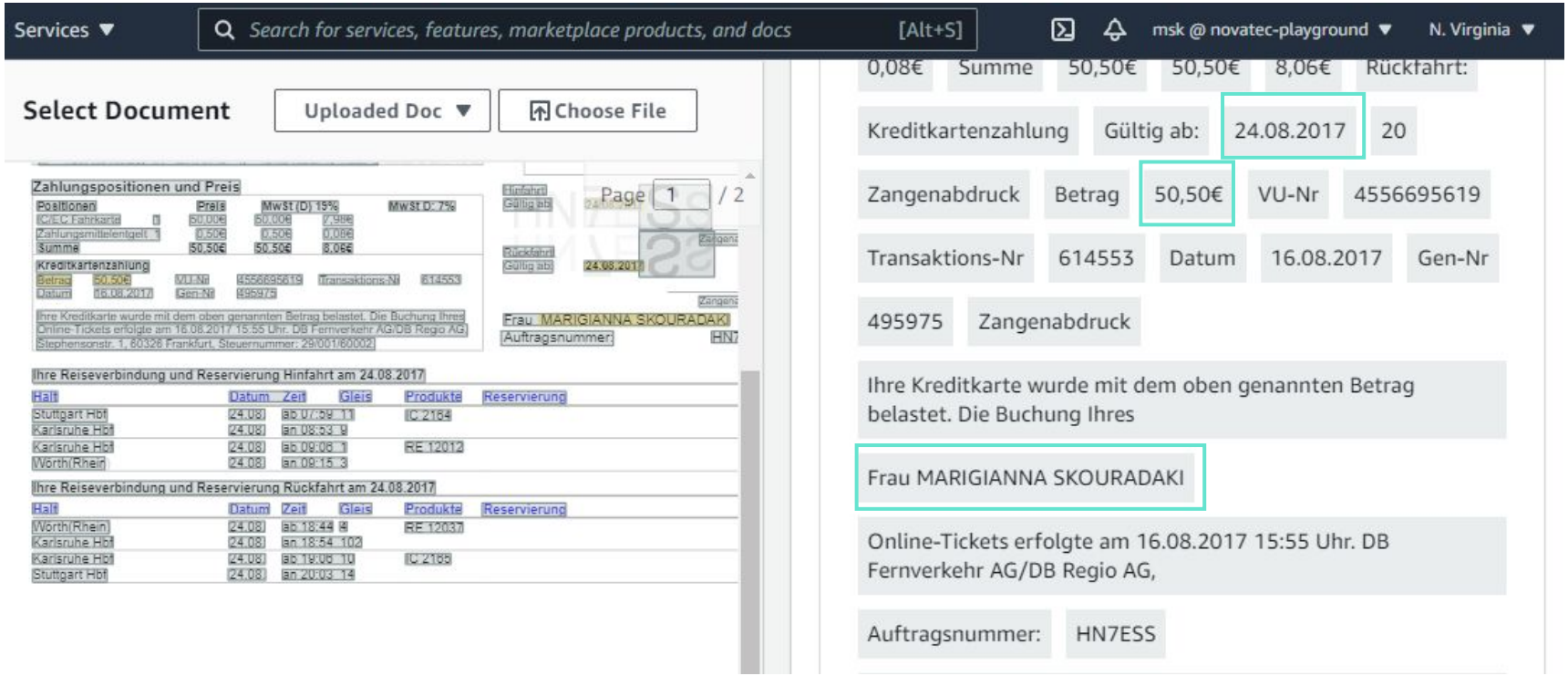
Intelligent Costs Submission Process

Semi-automated Cost-Submission Process



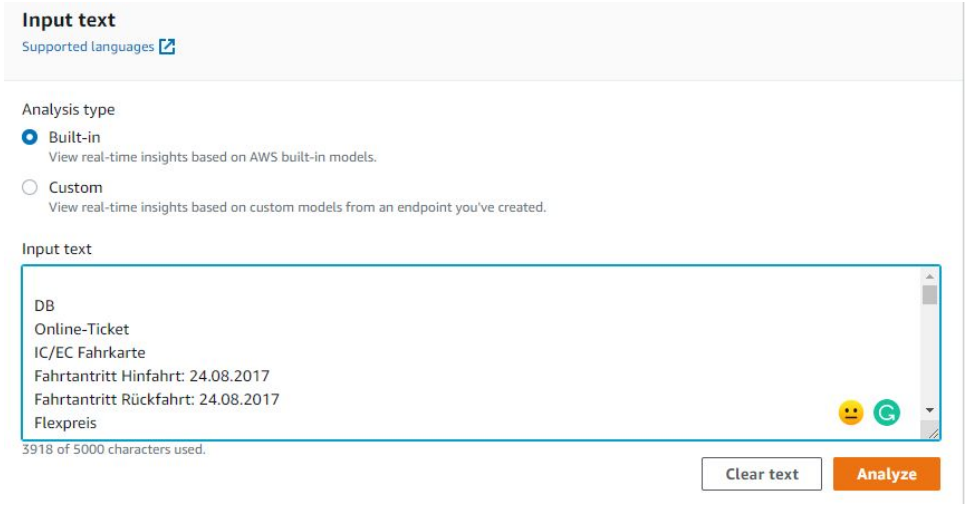
NOVATEC

AWS Textract (MLaaS)



NOVATEC

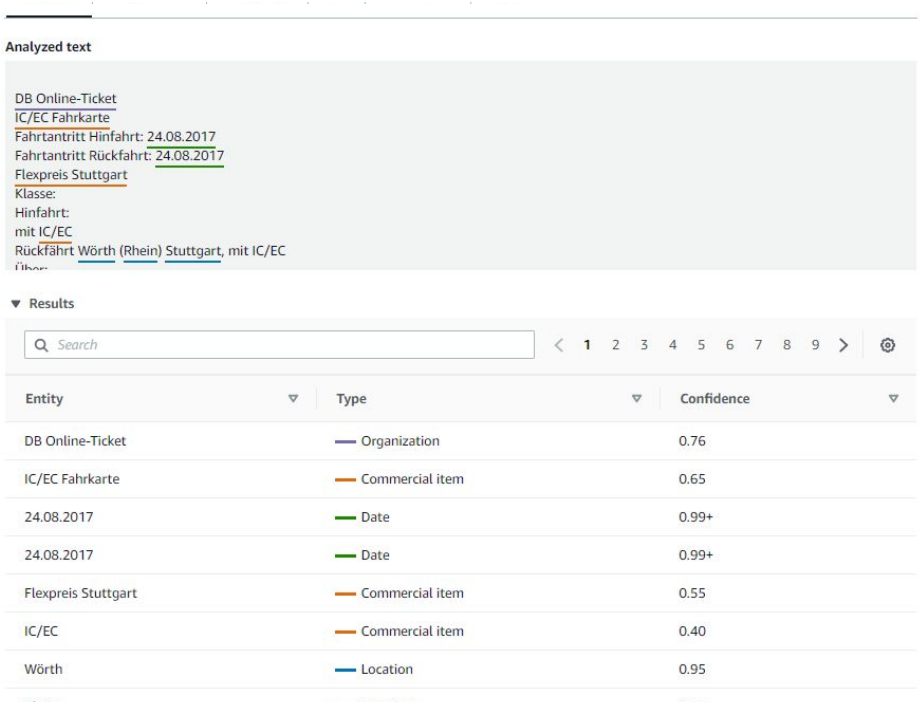
AWS Comprehend (MLaaS) 1/2



By default, Amazon Comprehend identifies entities:

- COMMERCIAL_ITEM
- DATE
- EVENT
- LOCATION
- ORGANIZATION
- OTHER
- PERSON
- QUANTITY
- TITLE

AWS Comprehend (MLaaS) 2/2



Analyzed text

DB Online-Ticket
IC/EC Fahrkarte
Fahrtantritt Hinfahrt: 24.08.2017
Fahrtantritt Rückfahrt: 24.08.2017
Flexpreis Stuttgart
Klasse:
Hinfahrt:
mit IC/EC
Rückfahrt Wörth (Rhein) Stuttgart, mit IC/EC

Results

Search

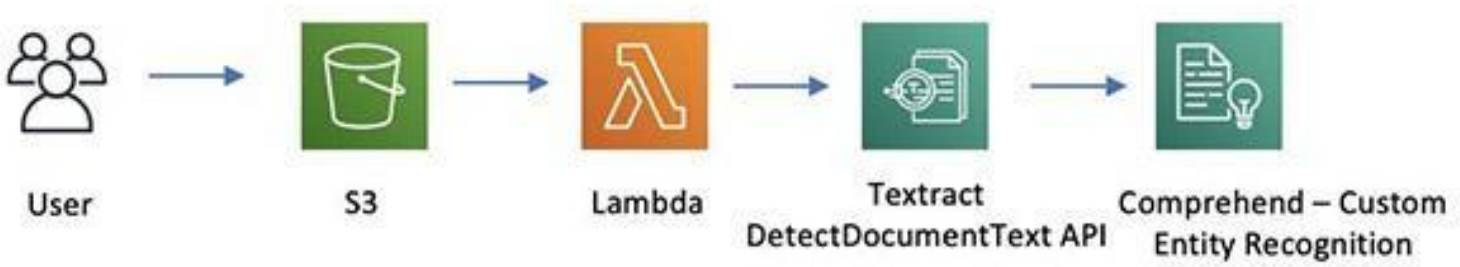
1 2 3 4 5 6 7 8 9

Entity	Type	Confidence
DB Online-Ticket	Organization	0.76
IC/EC Fahrkarte	Commercial item	0.65
24.08.2017	Date	0.99+
24.08.2017	Date	0.99+
Flexpreis Stuttgart	Commercial item	0.55
IC/EC	Commercial item	0.40
Wörth	Location	0.95

AWS Comprehend returns a **confidence score** and positional information of the detected entity

End-to-end Intelligent Document Analysis with AWS

- Async,
- 30s round-trip
- Replaced by Py2PDF (3rd Party Python Library)



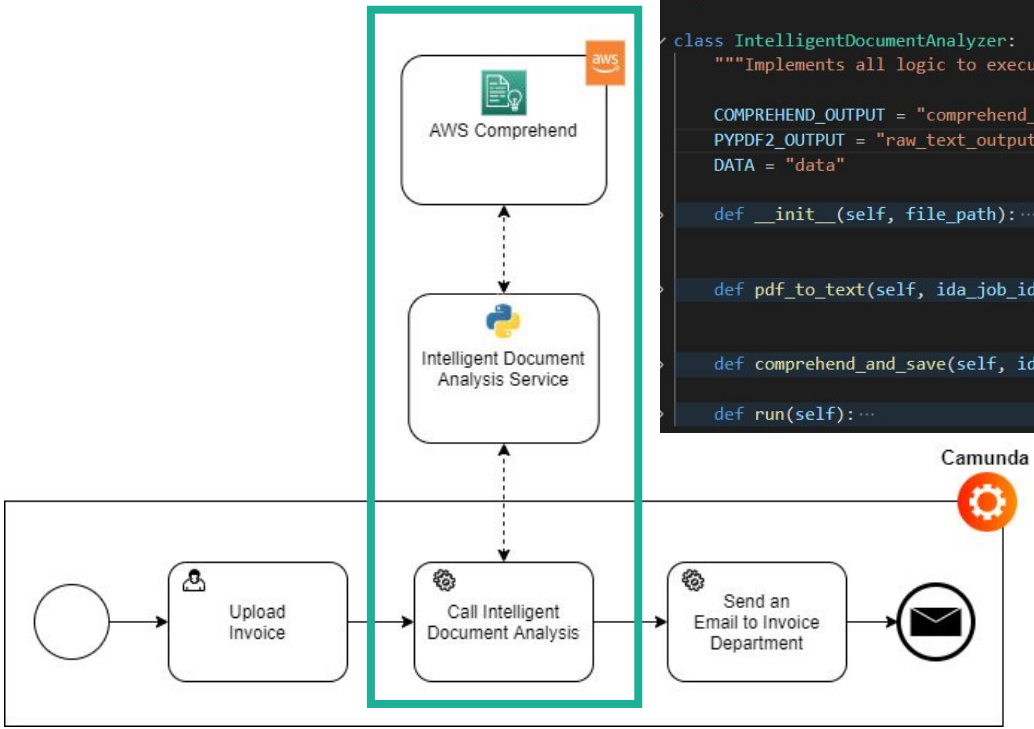
Source: https://aws.amazon.com/blogs/machine-learning/extracting-custom-entities-from-documents-with-amazon-textract-and-amazon-comprehend/

- Supports sync
- Async lasts 6min
- Custom training requires 10 documents per custom label

NOVATEC

Architecture

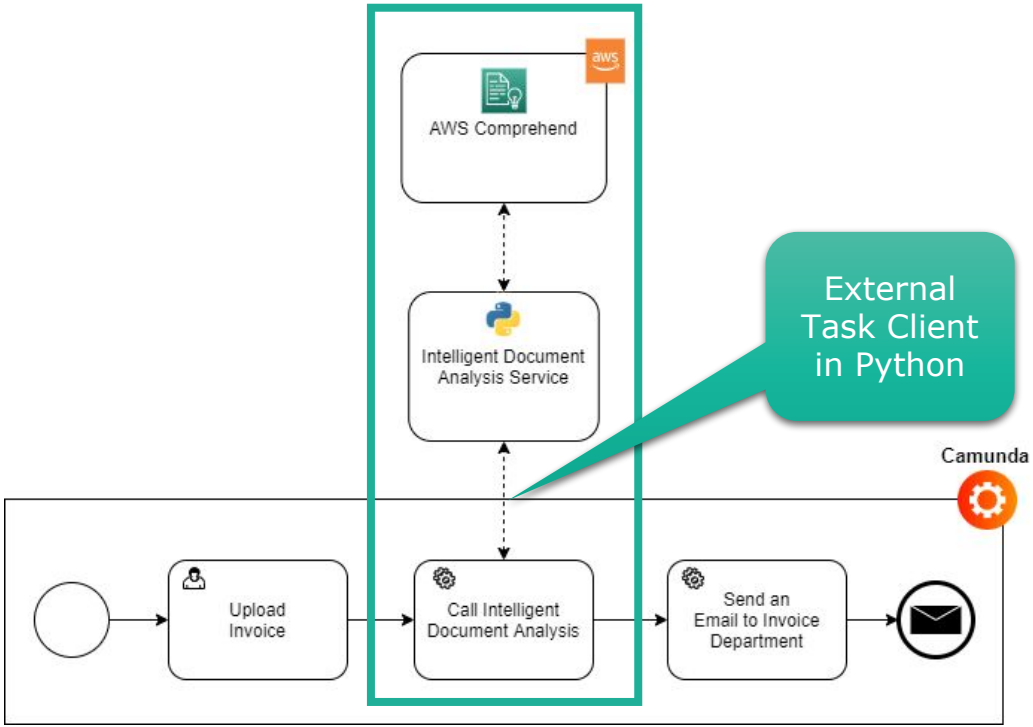
```python
import boto3
import json
import logging
import os
from os import path
import pandas as pd
import PyPDF2
import uuid

class IntelligentDocumentAnalyzer:
    """Implements all logic to execute Named Entity Recognition on a document."""

    COMPREHEND_OUTPUT = "comprehend_output"
    PYPDF2_OUTPUT = "raw_text_output"
    DATA = "data"

    def __init__(self, file_path):···


    def pdf_to_text(self, ida_job_id):···


    def comprehend_and_save(self, ida_job_id, raw_text_file_path):···

    def run(self):···
```



https://www.novatec-gmbh.de/en/blog/ipa-camunda-comprehend/

NOVATEC

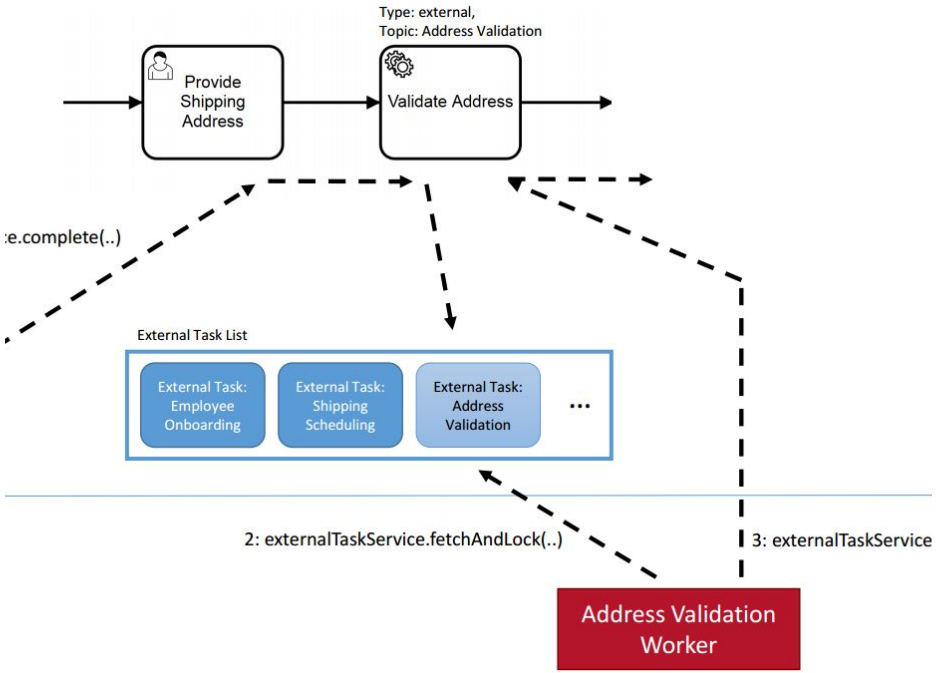
Implementation Details



https://www.novatec-gmbh.de/en/blog/ipa-camunda-comprehend/

NOVATEC

External Task Pattern



- External Task is a service task that provides a unit of work in a list that can be **polled** by workers
- **Polling:** no information is pushed by Camunda

Source: https://docs.camunda.org/manual/7.15/user-guide/process-engine/external-tasks/

External Task Client in Python

CamundaCon, 2019

```
AWS: Add Debug Configuration | AWS: Edit Debug Configuration (Beta)
def main():
    logging.info("Intelligent Document Analysis is running")
    while(True):
        camunda_client= camunda_external_task.client(PROCESS_ENGINE_URL, "ida_task")
        camunda_client.subscribe(topic = "IDA")

        file_path = get_invoice_file(PROCESS_ENGINE_URL, camunda_client.processInstanceId)

        extracted_values_dict = extract_data_from_file_to_dict(file_path)

        camunda_client.complete(**extracted_values_dict)

        logging.info("WorkerId: {} of process instance id: {} completed".format(camunda_client.workerid, camunda_client.processInstanceId))
    return
```

Alternative:
https://github.com/camunda-community-hub/camunda-external-task-client-python3

Chosen Camunda Deployments

Camunda Platform Run

Simple and Powerful

Java background not needed

Enables fast service orchestration

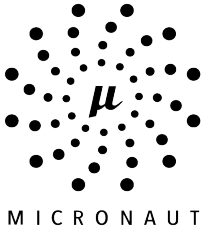
Camunda Micronaut

Developed in Novatec with the support of the open-source community

Lightweight

Faster start-up times

Less memory consumption



https://www.novatec-gmbh.de/en/blog/micronaut-meets-camunda-bpm/

NOVATEC

Demo

Takeaways

- IPA is about leveraging the automation level of your business process with intelligent services
- Now is the right time to apply IPA and it can be very profitable to the business
- It is relatively easy to connect AI logic to Camunda
  - No deep understanding of ML needed
  - MLaaS and External Task Client in Python open a lot of possibilities

Interested? Get in touch!

Blog Post
https://www.novatec-gmbh.de/en/blog/ipa-camunda-comprehend/

Intelligent Process Automation Offering
https://www.novatec-gmbh.de/beratung/intelligente-prozessautomatisierung/

MLOps Offering
https://www.novatec-gmbh.de/beratung/mlops/

Dr. Marigianna Skouradaki
Consultant
marigianna.skouradaki@novatec-gmbh.de

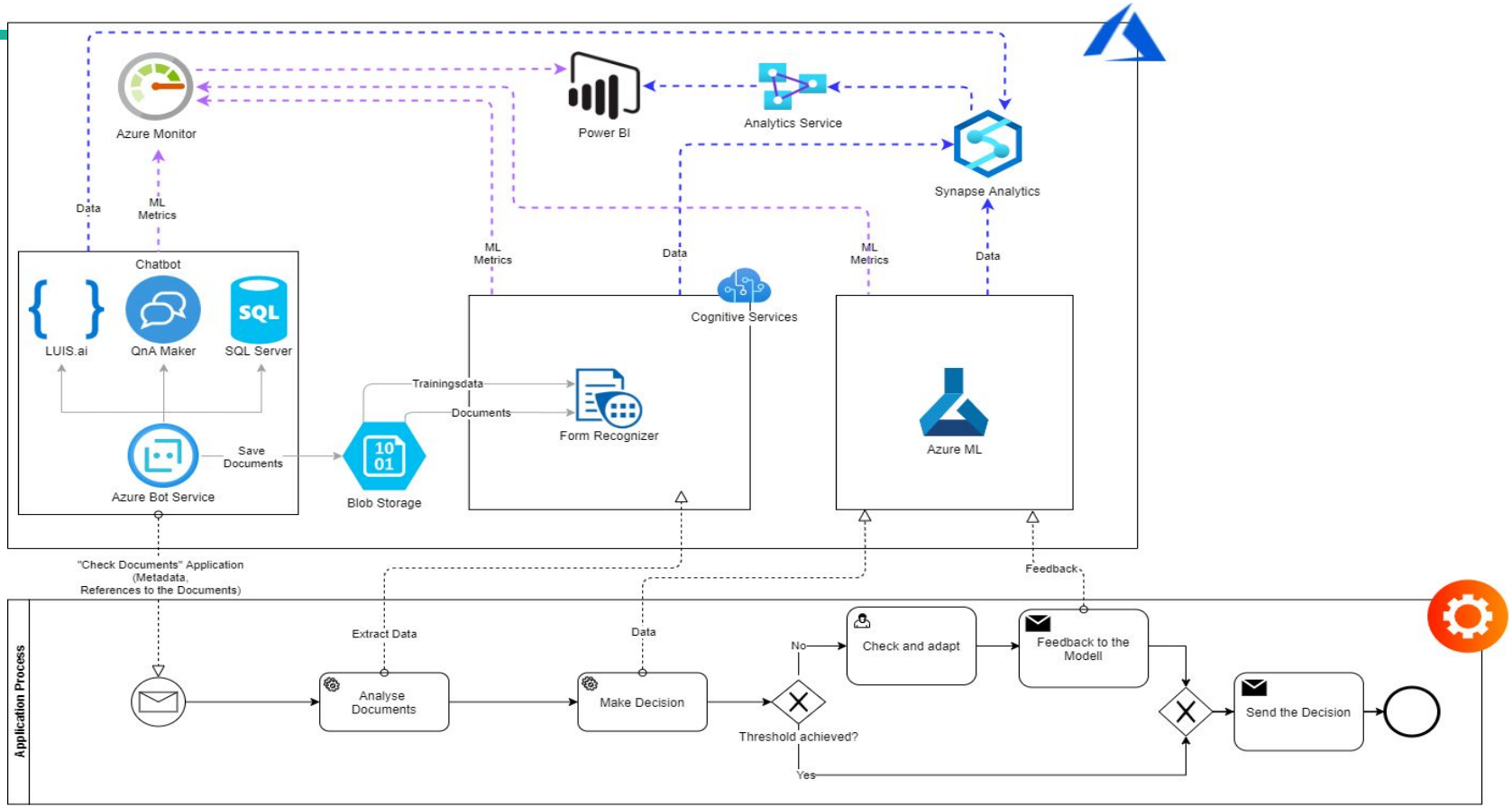
NOVATEC

Quiz

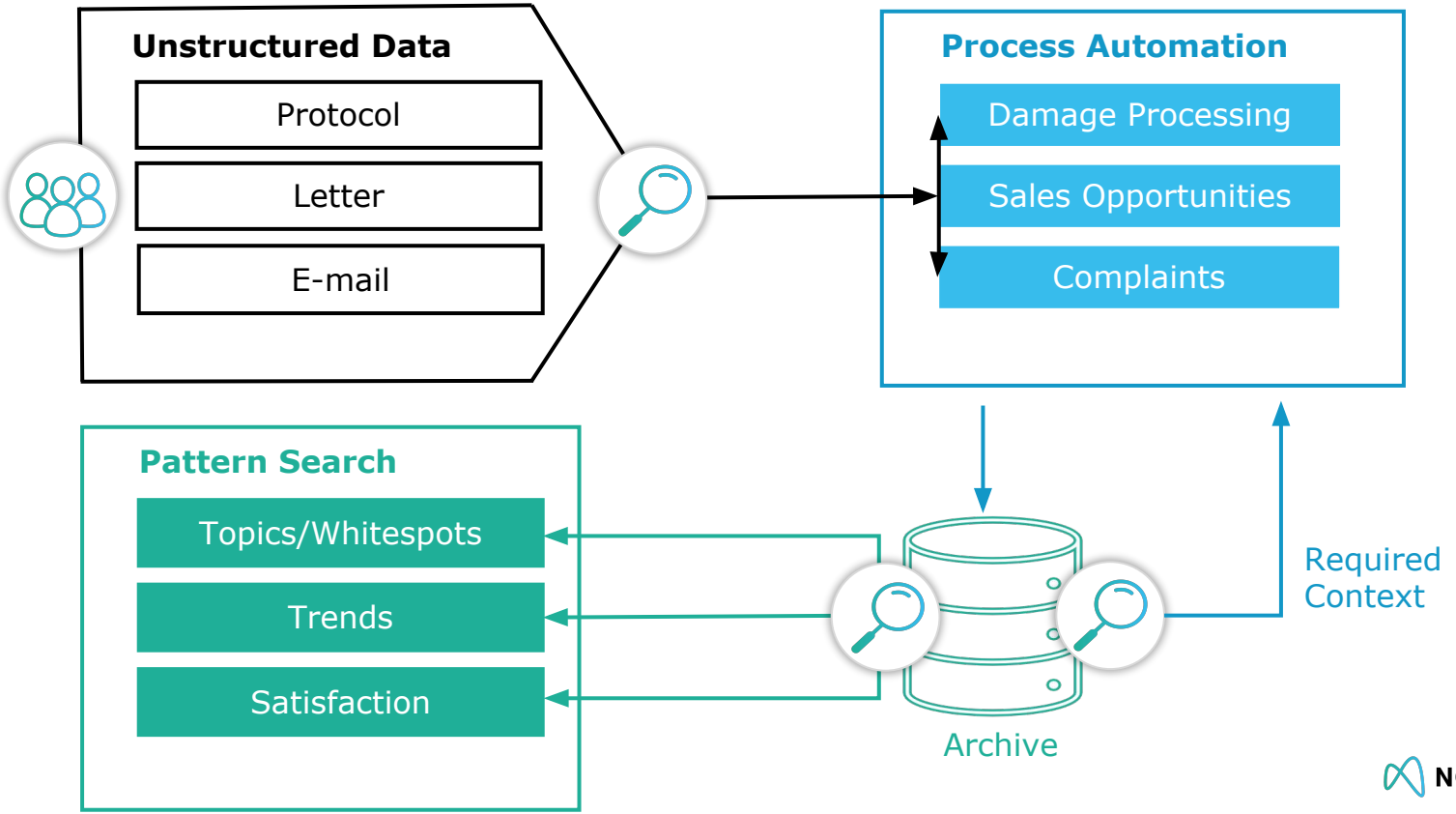
https://www.menti.com/vbjnorxu7f
Code: **6828 4009**

Which terms are related to Intelligent Process Automation?

Example: Azure Architecture

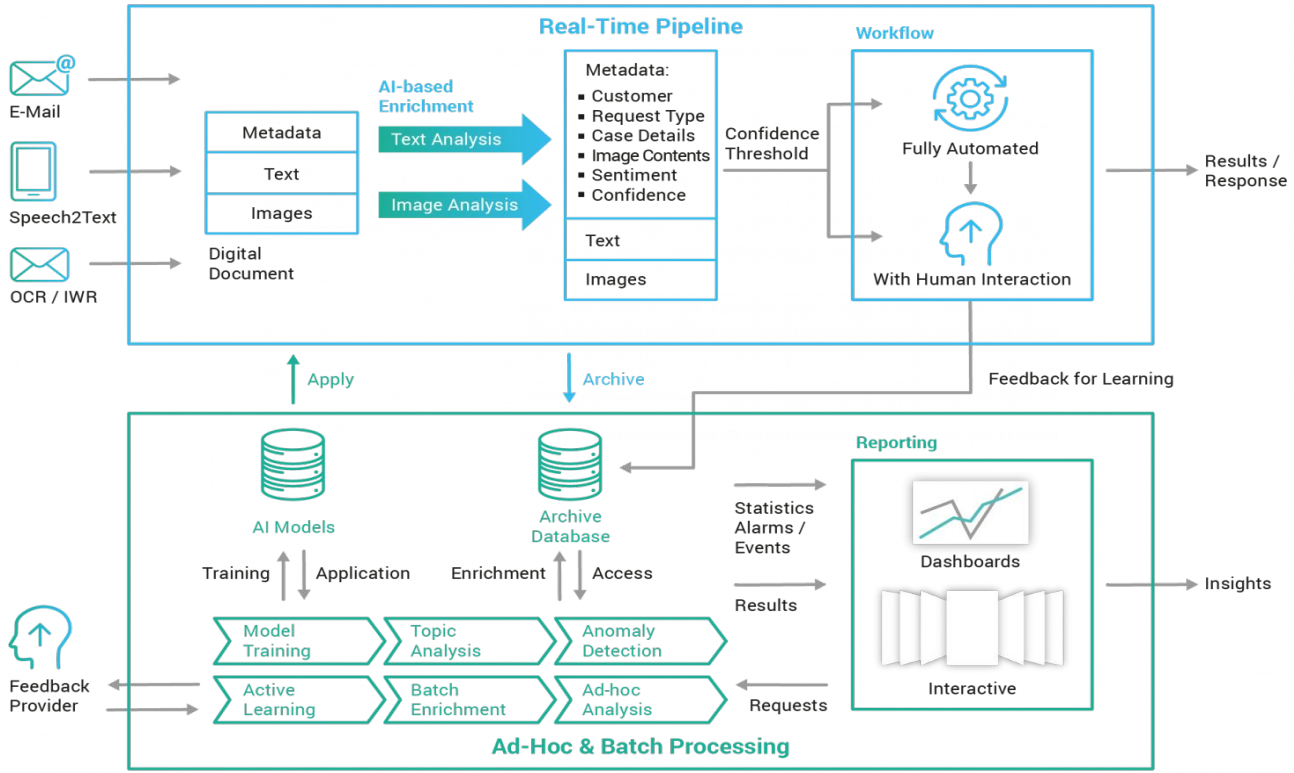


Intelligent Document Analysis in Action



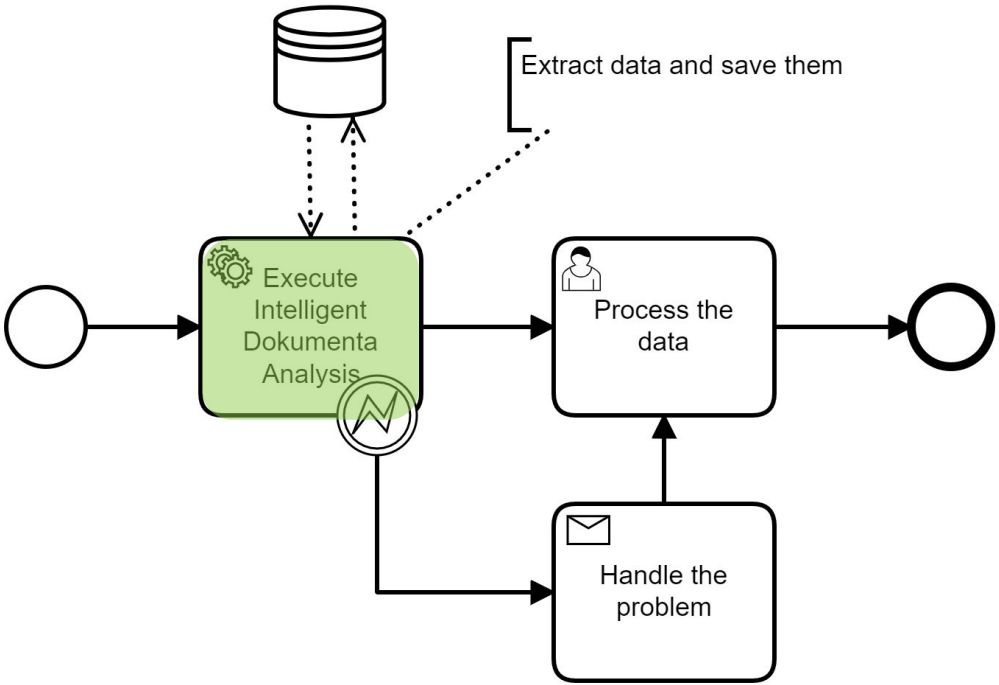
NOVATEC

Functionality of Intelligent Document Analysis



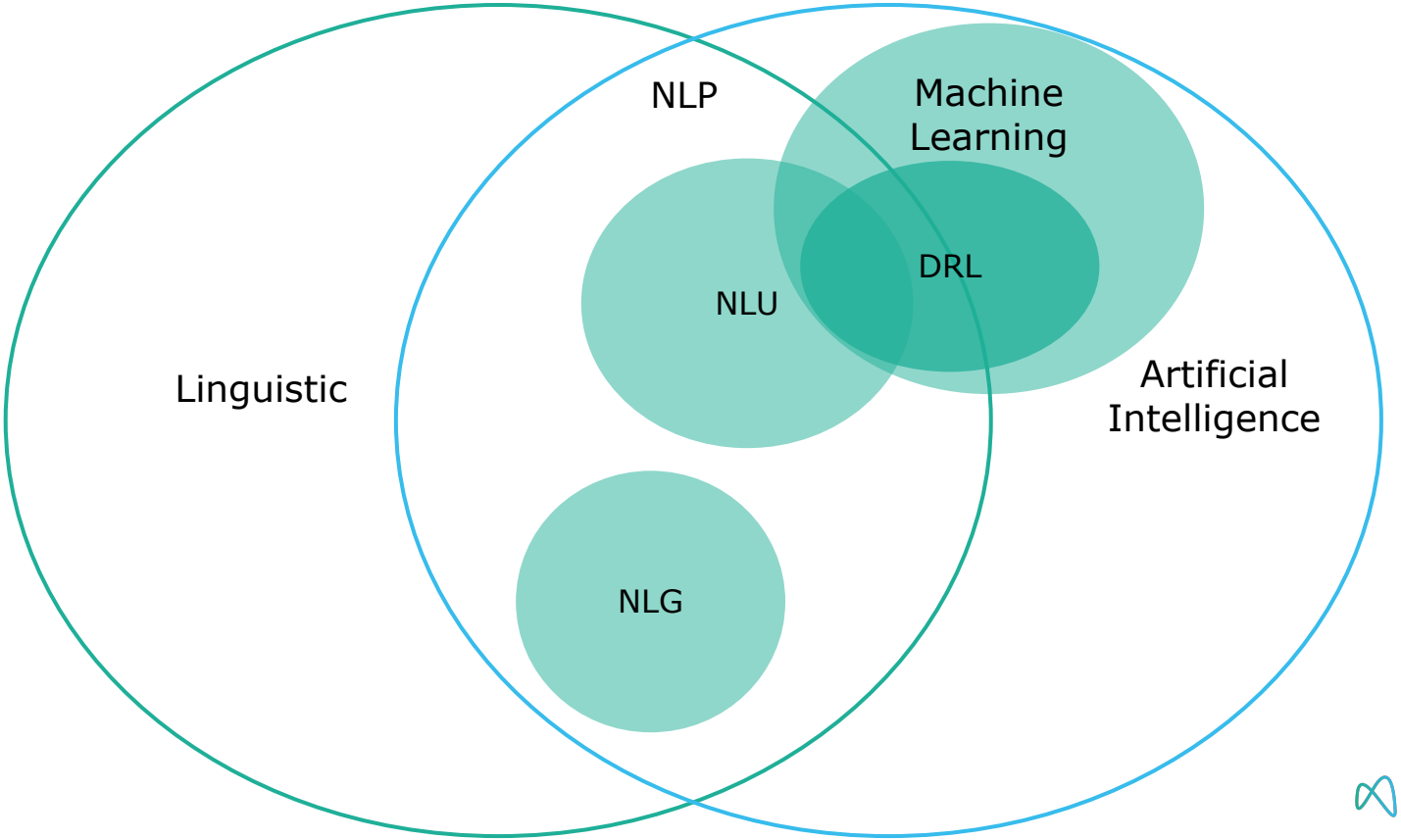
NOVATEC

Integration of IDA in BPM



NOVATEC

Natural Language Processing (NLP)



NOVATEC

Limitations

Textract

- Supports only asynchronous detection of text in PDF files
- Response times for analyzing one file were approximately 30s round trip
- To achieve better response times, we replaced Textract functionality with Py2PDF local library

Comprehend

- Custom entity recognition requires 10 documents per custom label
- Asynchronous call with one PDF file as input takes approximately 6 minutes to respond
- Although not tested in detail, we suspect that this time would remain stable even with a larger batch of files.

Intelligent Document analysis

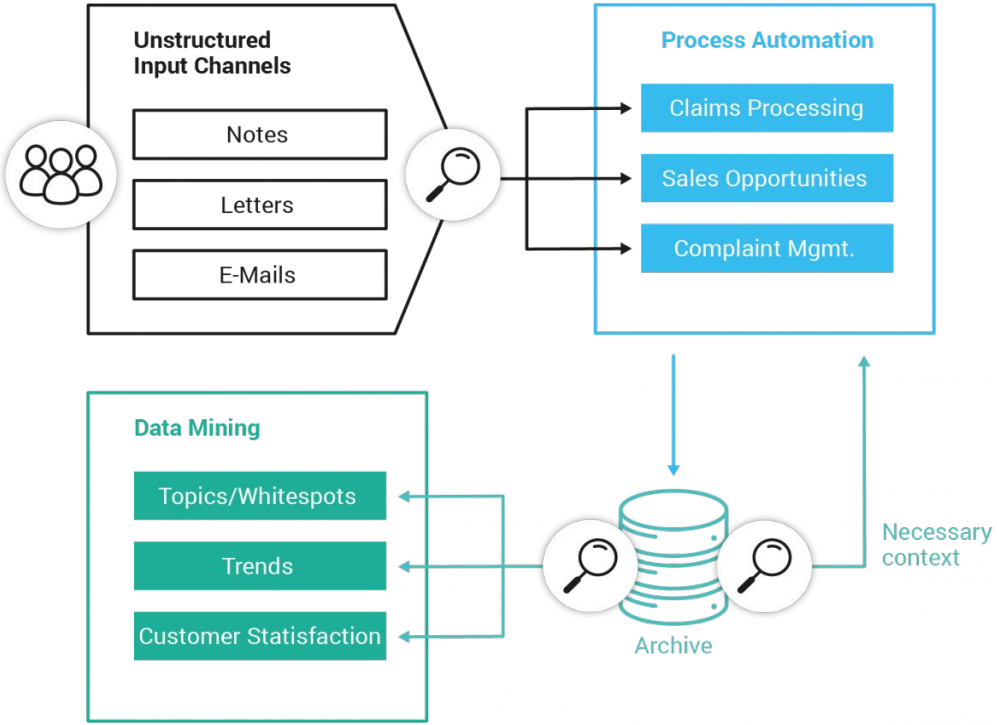
Problems with current methods of analysis

- Big data: Too many documents that are impossible to analyze manually.
- Analysis processes that are slow because they are not automated
- Knowledge loss because it was not possible to process all the information

Benefits as a result of intelligent document analysis

- Customer satisfaction increased as a result of faster processes
- Improved quality of services as a result of greater know-how
- (Product) improvements as a result of pattern recognition and trend recognition in documents
- Targeted, personalized and sustainable addressing of customers as a result of data-driven marketing

Intelligent Document Analysis in Action



NOVATEC

MLOps: DevOps for Machine Learning

Problem definition for manual training & operations

- Long roll-out cycles for new models due to a manual training process
- No continuous and standardized quality assurance of data and models
- No automation, standardization and transparency can be achieved due to lack of data, experiment and model management

Benefits of MLOps

- Automation through standardized ML pipelines (e.g., training & model deployment)
- Bridging the gap between development phase and practical deployment
- Versioning of data, hyperparameters & models
- Data and model testing to identify issues as early as possible in the ML pipeline
- Production monitoring to ensure models are performing to expected quality as they process new data

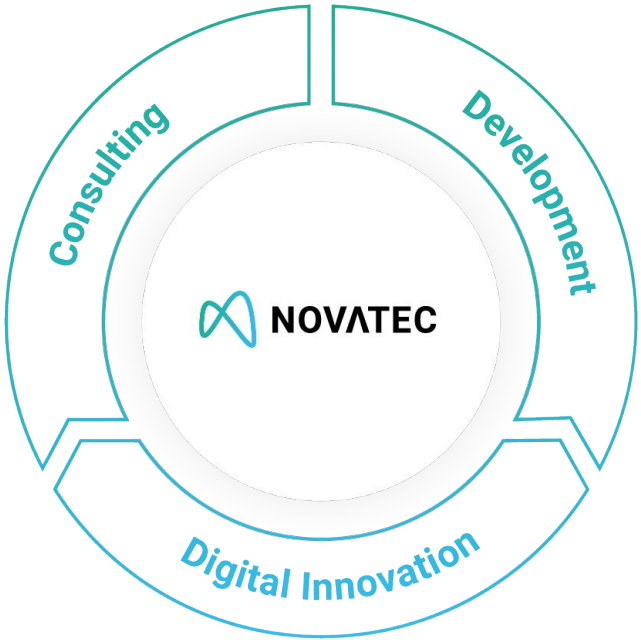
Source: https://ml-ops.org/content/motivation
Source: https://www.kdnuggets.com/2020/12/mlops-why-required-what-is.html

Translate Recognized Entities to own Model

```
class EntitiesExtractor:
    """Contains all logic to extract relevant information from the recognized entities and save them t
    It currently works for DB Tickets.
    """

    def __init__(self, entities_file_path):···

    def extract_organization(self):···

    def extract_name(self):
        """Extracts the full name (firstname, lastname) from the recognized entities file."""

        globals().update(self.__dict__)

        name_fields = data[(data["Type"] == "PERSON")]
        if(name_fields.shape[0] != 0):
            name=""
            for field in name_fields.itertuples():
                    name = name + " " +field.Text

            names = name.split(" ")[:-1]
            extracted_data["firstname"] = ' '.join(names)
            extracted_data["lastname"] = name.split(" ")[-1]
        else:
            extracted_data["firstname"] = np.nan
            extracted_data["lastname"] = np.nan
```


NOVATEC

Accompanying Novatec into the digital future



Our Portfolio

As an independent IT specialist, we have been leading our customers into the digital future since 1996.

Novatec in Numbers

9
Locations

250
Employees

35
Ø Age of employees

29 Mio. €
Revenue in 2019

NOVATEC

Our Portfolio

Agile Organization

Quality Engineering

Agile Software Engineering

IT Architecture & Cloud

Enterprise Architecture Management

Artificial Intelligence & Machine Learning

Internet of Things (IoT)

Augmented & Virtual Reality

Business Process Management

CAMUNDA | Platinum Partner CERTIFIED

Application Performance Management

NOVATEC