IFC-Bank of Italy Workshop on "Data Science in Central Banking: Applications and tools"

14-17 February 2022

Natural language processing for risk management[1]

Bijan Sahamie,
Deutsche Bundesbank

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

Natural Language Processing for Risk Management

Discussion of Use-Cases

Dr. Bijan Sahamie, Deutsche Bundesbank

Disclaimer

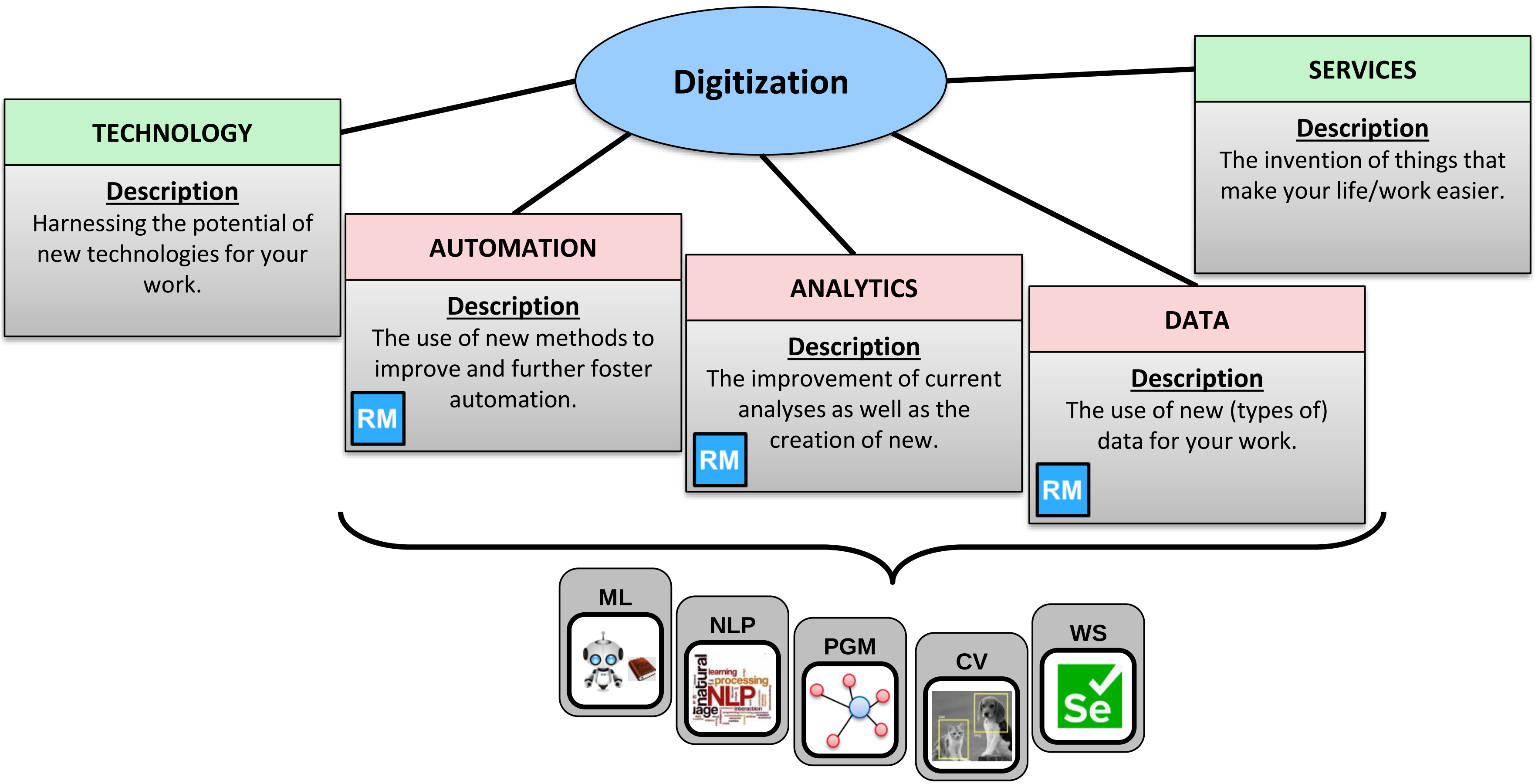
The opinions, views, facts, interpretations expressed within the confines of these slides are those of the speaker alone and have nothing to do with Deutsche Bundesbank. All opinions, views, facts, interpretations expressed by the speaker orally during the talk, before, or thereafter are explicitly those of the speaker himself and are not in any way connected with Deutsche Bundesbank.

All mistakes, bad impressions, all weirdness or non-professionalism you perceive either from these slides or the speaker do not in any way represent Deutsche Bundesbank and are the sole responsibility of the speaker.

Let‘s have a very high-level overview

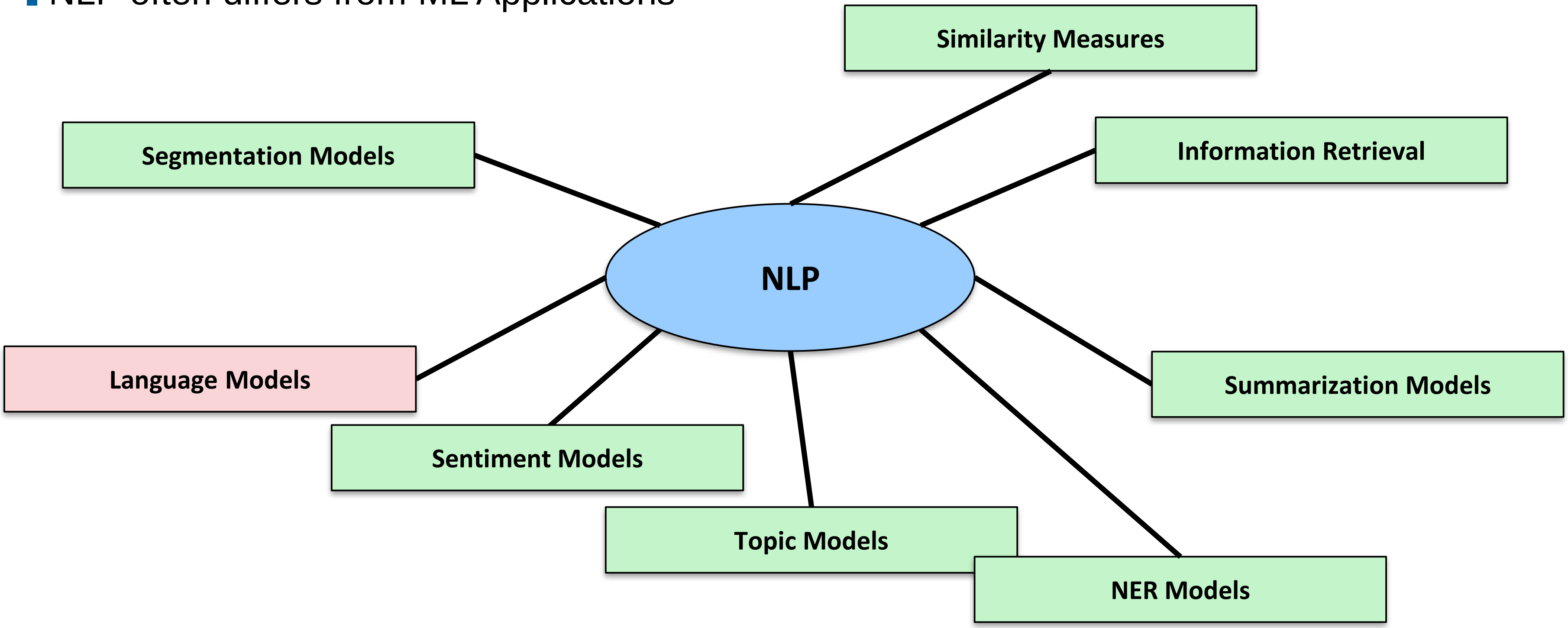
What are the main areas of interest and which technologies might be a good choice?

Let us have a look at the main areas of Digitization



Natural Language Processing

NLP often differs from ML Applications



Two Properties of NLP that stand out.

- Your use-case, although it might be specialized, is an (immediate) derivative of a general purpose task. Both are intimately related.
- Inside your IT-system there are several intertwined machine learning components. The machine learning system might even work as a hierarchy.

Counterparty Risk
Use-Cases in Risk Control

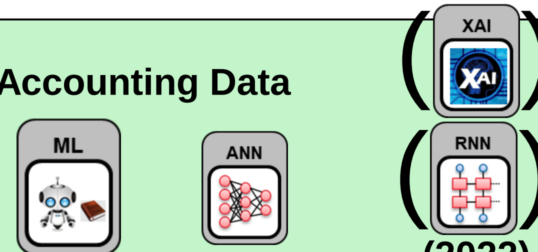
There are use-cases for both machine learning and natural language processing in Risk Control

Prediction of Financial Distress using Accounting Data

<u>What's it about:</u> „Classic" forecasting using machine learning

<u>Primary Utility:</u> Generating a short-list for our analysts

<u>Models:</u> Ensemble of Neural Nets with calibrated thresholds (in the sense of a Outlier Detection)



(2022)

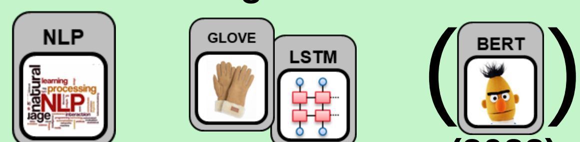
ONLINE

(Fact-associated) Financial Sentiment of News Messages

<u>What's it about:</u> Automated assessment of news messages

<u>Primary Utility :</u> Workload reduction (i.e. improved efficiency)

<u>Modelle:</u> Mixture of Experts – Approach incl. LSTMs + Glove



(2022)

ONLINE

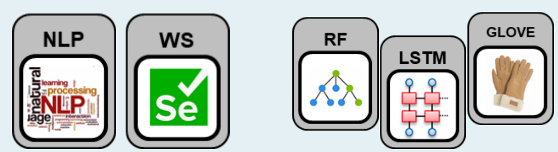
(Doing webscraping since nov 2019)

Filtering of News Messages sourced through webscraping

<u>What's it about:</u> Filter out uninteresting messages

<u>Primary Utility:</u> Better signal/noise-ratio, scalability of the approach

<u>Models:</u> (Self-Defined) Frequency Measures, Random Forests, LSTMs and/or GRUs + Glove.



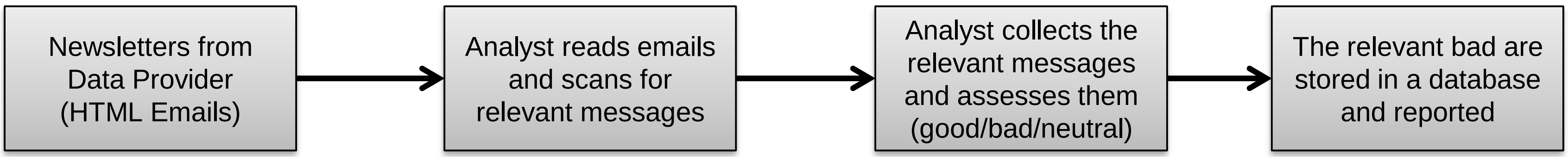
WORK IN PROGRESS

(2022)

Financial Sentiment Analysis of News Data

Overview

Imagine the process. (very rough and simplified sketch, not entirely correct for the sake of brevity.)



Question. Can we make this more efficent?

Goal.

- Let a machine read the messages and just present to the analyst a pre-selected list (based on the system output)

Cost/Benefit compromise.

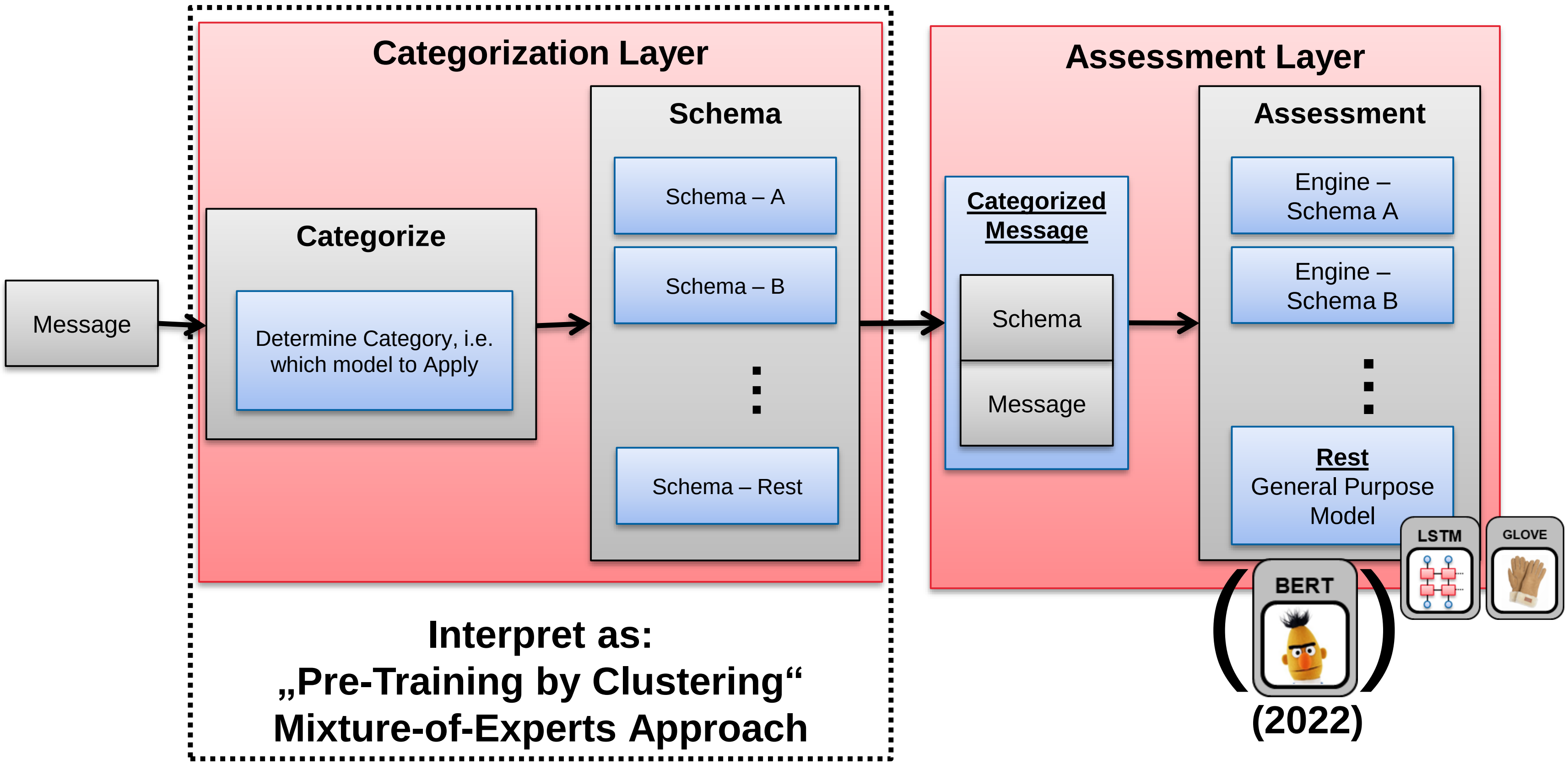
- Potential benefit: You lower the workload for the analyst
- Potential benefit: You raise the consistency of the process output
- Potential cost: You might loose some of the interesing messages.

ML-Dev Task.

- Find a sensible sweet-spot in the Cost/Benefit regime and optimize your system towards it

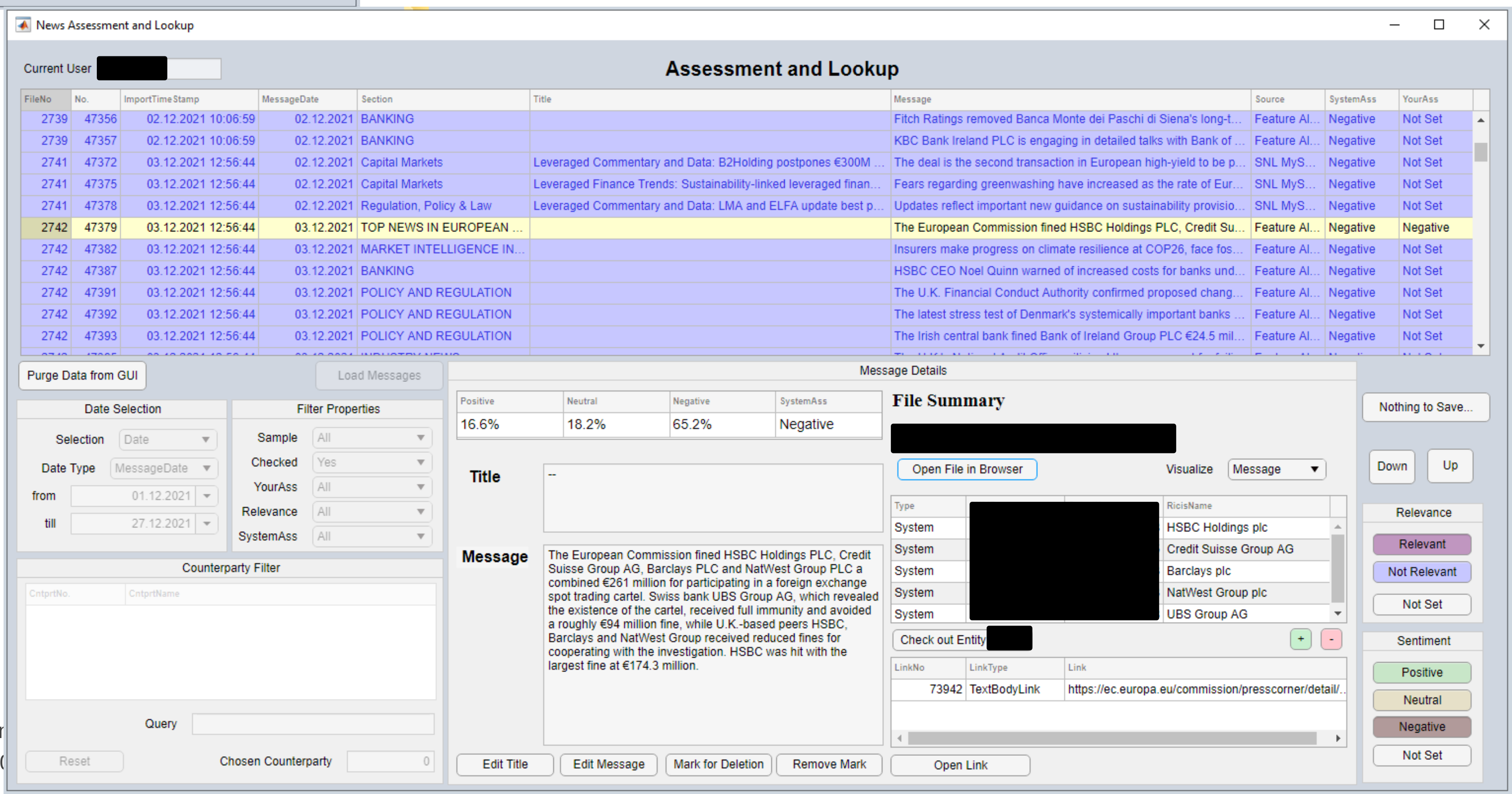
Financial Sentiment Analysis of News Data

The Model Concept



Financial Sentiment Analysis of News Data

A Screenshot



Filtering Techniques for Webscraping of News

Filtering Techniques for Webscraping of News

Volume per Run (currently):

- ~15000 search results are scanned
- ~900 website jumps

Data Gathering - Phase

Google-News Searches

Extract Information From Search Results

Filtering - Phase

Filter Out News that are not interesting

Go to News Source and Search for Showstoppers

Delivarable (to Analyst).

Produce Formatted Excel List

Date
Source
Title
Summary
Link to Source

ML

At the moment, these filters are really simple

The goal is to replace the simple filters with ML-driven filters

Questions. Is it possible to…

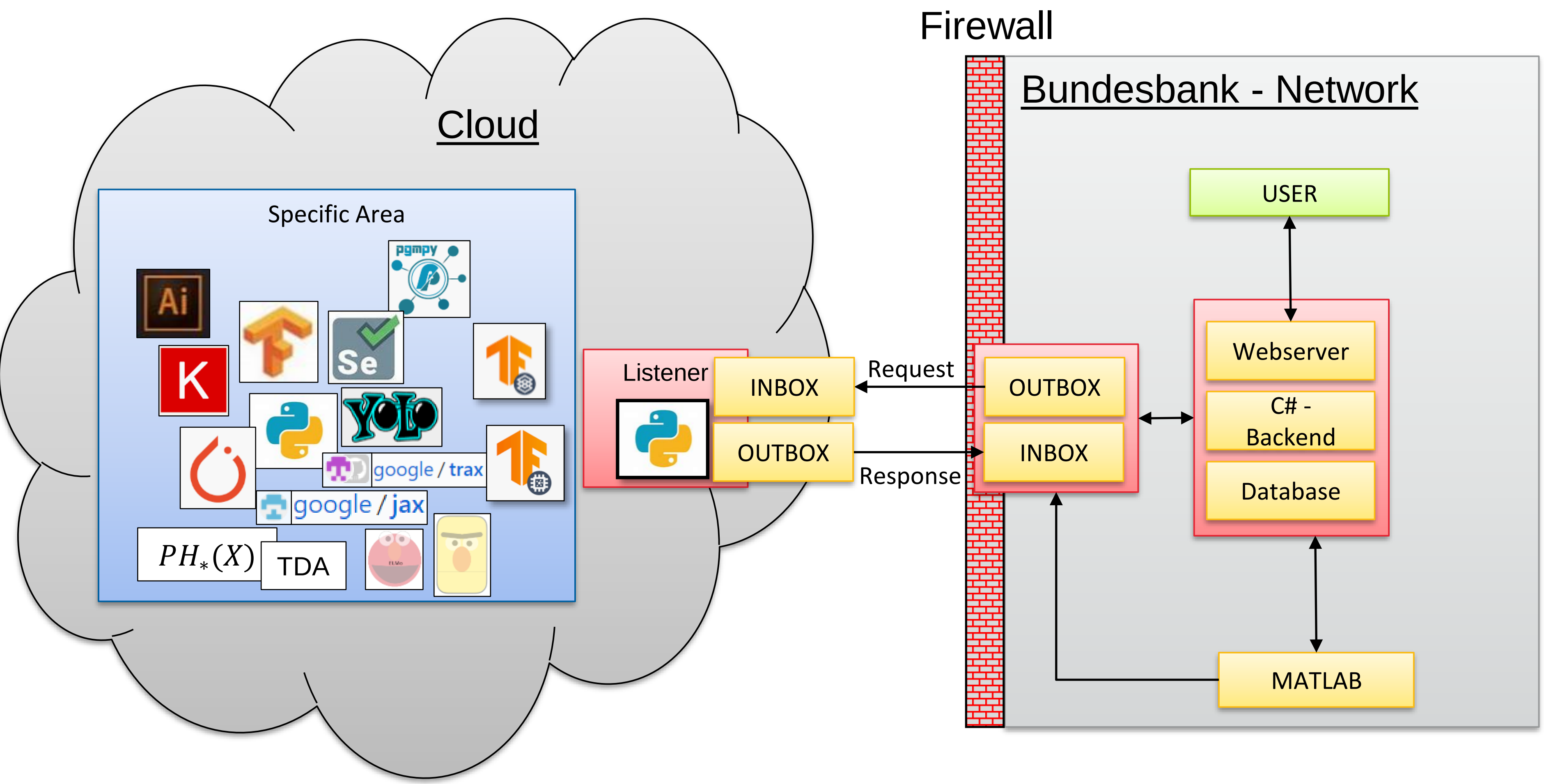
ML

- … distinguish „useful“ from „non-useful“ sources by just looking at their name, and is this a learnable task?
- … distinguish „useful“ from“ „non-useful“ sources by using frequency-measures?
- … distinugish „useful“ from „non-useful“ messages by looking at the titles?
- … distinugish „useful“ from „non-useful“ messages by looking at the summary?

Answer.
YES!

Filtering Techniques for Webscraping of News

(Target) IT – Architechture



The End