Image: used under license from shutterstock.com

Data enrichment and model creation using text mining and other unstructured data

Dr. Andreas Nawroth
Head of Artificial Intelligence
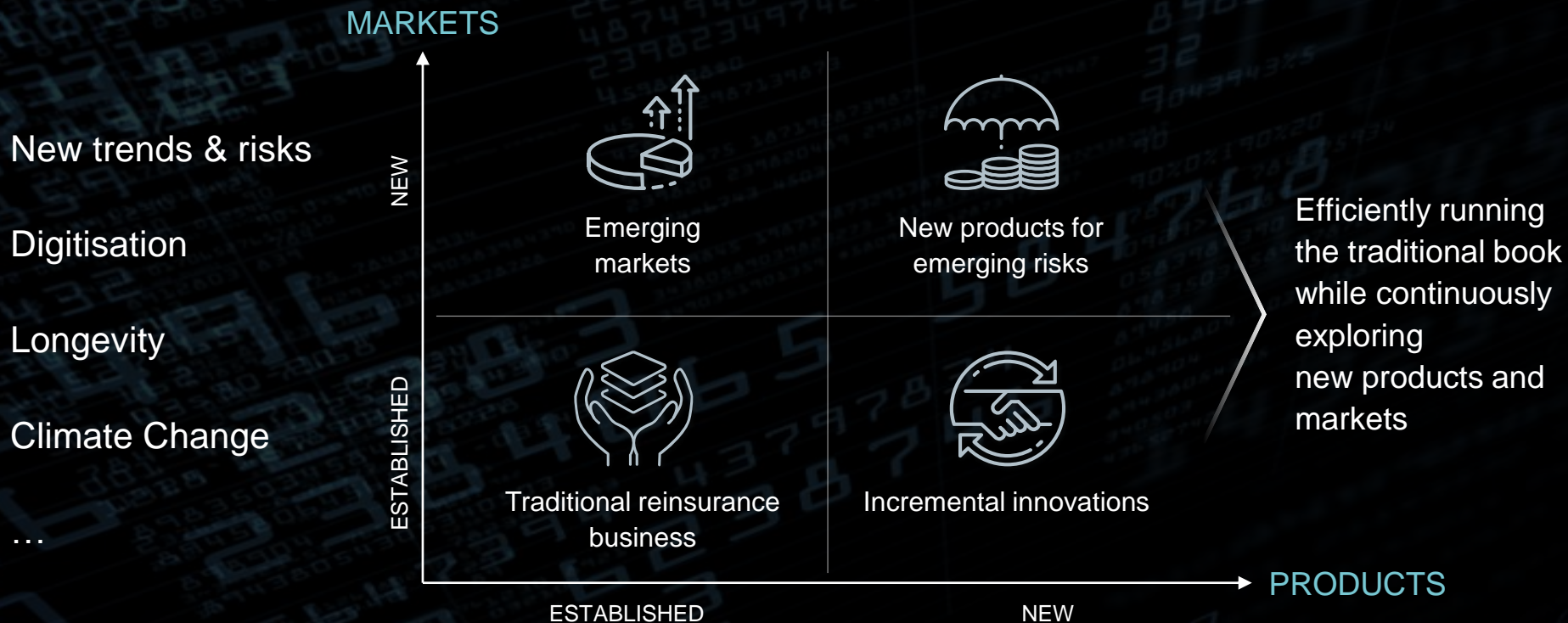German Data Science Days 2019
19. Feb 2019

Munich RE

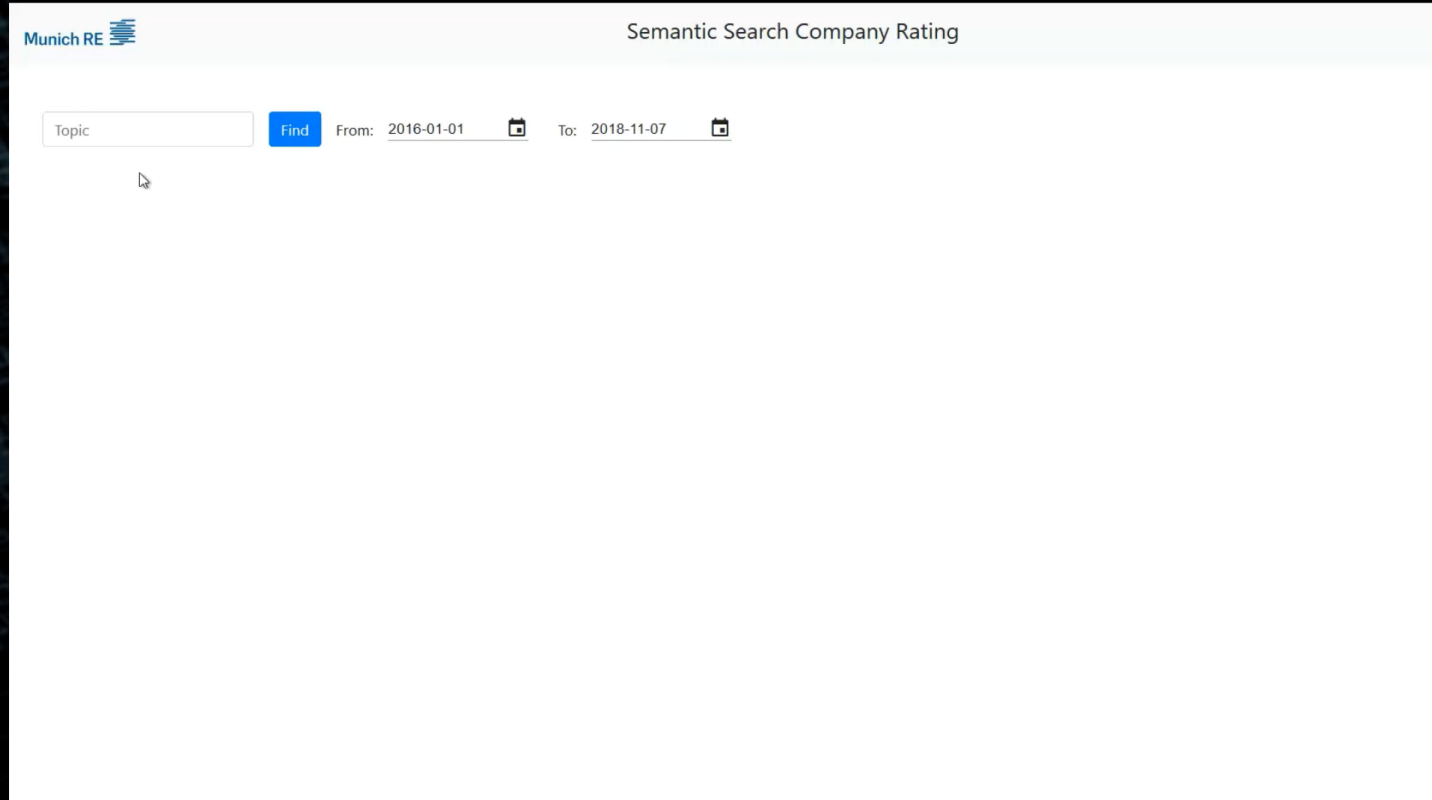# Agenda for data enrichment and model creation using text mining and other unstructured data

Munich RE

1 Introduction to Analytics & Artificial Intelligence

Munich Re actively shapes the transformation of the (re-)insurance industry

MARKETS

New trends & risks

Digitisation

Longevity

Climate Change

…

	ESTABLISHED	NEW
NEW	Emerging markets	New products for emerging risks
ESTABLISHED	Traditional reinsurance business	Incremental innovations

PRODUCTS

Efficiently running the traditional book while continuously exploring new products and markets

Example: Ad-hoc risk identification and quantification

Example: Extraction of timeline of events

In recent years, there has been many applications of text mining

Web Crawling

Fire Detection

Chat Bot

Customer Satisfaction
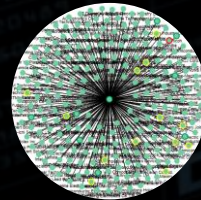
High-Tech Supply Chain

Claims Classification
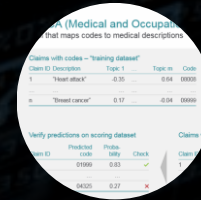
Medical Data
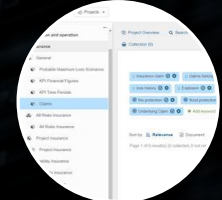
Investment Decisions

Oil & Gas Supply Chain

Investment Process
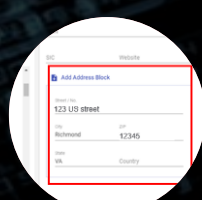
Address extraction
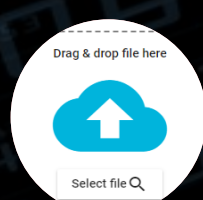
Data Lake
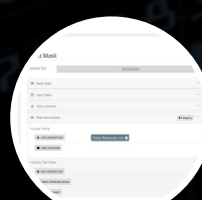
Structuring of Claims Data

Fact Extraction
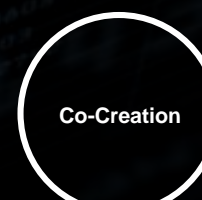
Co-Creation

2 **Deep dive into Deep Learning approach**

What is Text Mining: unlock the value of data in unstructured files

Unstructured data as input

Text Mining

Structured data as output

Text Pre-processing	Identify information	Deliver result
▪ OCR ▪ Tokenization ▪ Word Embeddings	▪ Named Entities ▪ Relation ▪ Sentiment	▪ Structured data in predefined format ▪ Risk Indices
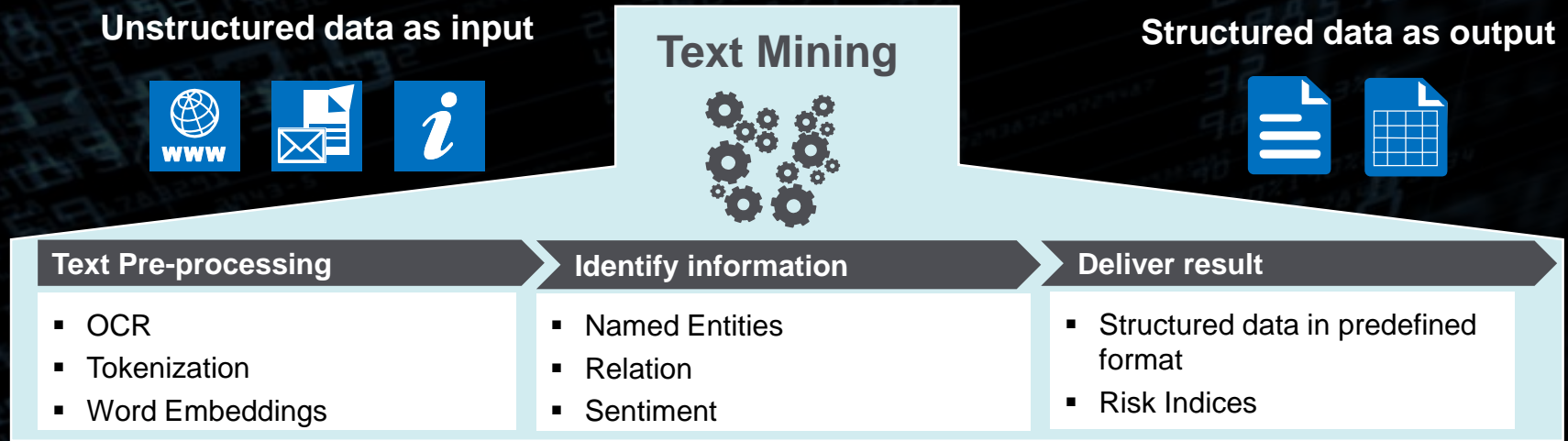
Evolution of approaches

Look-up methods → Rule-based approach → "Traditional" Machine Learning approach → Deep Learning approach

One most used text mining techniques: Named Entity Recognition (NER), solves "Who exactly, When exactly, What exactly"?

Munich RE

Illustrative example

Munich Re is expecting a loss of €1.4bn due to Hurricanes Harvey, Irma and Maria, for the third-quarter in 2017.

Desired outcome

Who: Munich Re

What: loss of €1.4bn

When: third-quarter in 2017
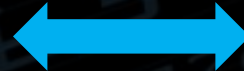
Data enrichment with additional data source (wiki, etc.)

NER challenge: how to extract the right info with the consideration of context?

With Munich Re's MIRA Digital Suite, life insurers are massively reducing the effort and expense involved in applications and claims. CLARA, for example, halves the average time taken to settle disability claims.

Republican lawmakers still think Google is biased against conservatives, Google still claims that it's not. The news agency reports.

https://www.healthinsuranceproviders.com/what-is-a-health-insurance-claim/

https://www.recode.net/2018/12/11/18136453/google-youtube-bias-sundar-pichai-testimony-congress
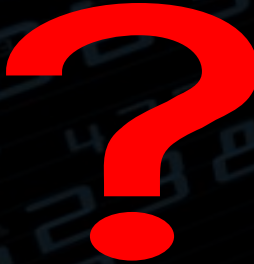
Fast innovations of text mining enable faster and more efficient info extraction for enhanced data quality and process automation

Milestones of text mining

2008 — NLP (almost) from scratch, Multi-task learning[1]

2013 — Word embedding (faster)

2013 — Neural networks for NLP

2014 — Sequence-to-sequence framework

2015 — Attention

2015 — Memory-based networks

2018 — Pre-trained language models

Relations of words captured by word2vec (Mikolov et al., 2013 a, b, c)

Male-Female   Verb tense

A convolutional neural network for text (Kim, 2014)
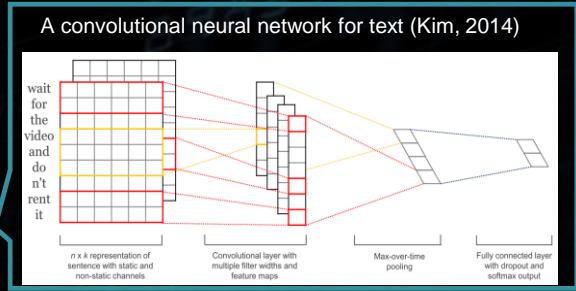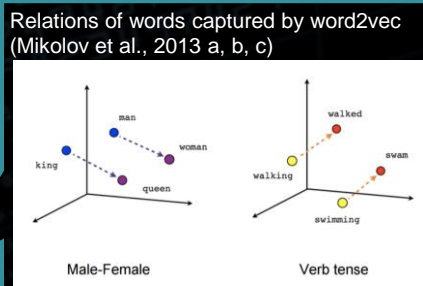
Enhanced efficiency: With this framework, Google in year 2016 started to replace its monolithic phrase-based machine translation (MT) models with neural MT models (Wu et al., 2016), replacing 500,000 lines of phrase-based MT code with a 500-line neural network model.[2]

Reduced limitation: Enables learning with significantly less data, only require unlabeled data.

- Mikolov et al. 2013a, b. c. Efficient Estimation of Word Representations in Vector Space. Distributed Representation of Words and Phrases and their Compositionality. Linguistic regularities in continuous space word representations.
- Kim, 2014. Convolutional Neural Network for Sentence Classification.
- Wu et al. 2016. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.
- 1. Collobert and Weston, 2008, A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning
- 2. https://www.oreilly.com/ideas/what-machine-learning-means-for-software-development

TM framework captures recent AI innovations – text mining processes become more similar to human brain processes

Models evolved from rule-based to those represent human brains*

U.T. Austin's Lonestar Advanced Computer Cluster

Blue Gene

IBM Watson

Human brain

1 PB

Rat brain

Memory

1 TB

Playstation 4

iPad 4

1 GB

Rat neocortical column

Single cellular model

1 Gigaflop   1 Teraflop   1 Petaflop   1 Exaflop

Computational speed

Deep learning neural network applied in text mining (illustrative)

Text

I   saw   Einstein   in   Monaco

TM framework captures recent AI innovations – text mining processes become more similar to human brain processes

Models evolved from rule-based to those represent human brains*

Memory

- 1 PB
- 1 TB
- 1 GB

U.T. Austin's Lonestar Advanced Computer Cluster

Blue Gene

IBM Watson

Playstation 4

iPad 4

Single cellular model

Rat neocortical column

Rat brain

Human brain

1 Gigaflop — 1 Teraflop — 1 Petaflop — 1 Exaflop

Computational speed

Deep learning neural network applied in text mining (illustrative)

Character-level CNN (Convolutional Neural Network)

+

Word-level embedding (Glove / Word2vec)

Text

C W

CNN

Glove / Word2vec

I   saw   Einstein   in   Monaco

* Based on illustration on wired.com

TM framework captures recent AI innovations – text mining processes become more similar to human brain processes

Models evolved from rule-based to those represent human brains*

Memory

1 PB

1 TB

1 GB

U.T. Austin's Lonestar Advanced Computer Cluster

Blue Gene

IBM Watson

Playstation 4

iPad 4

Single cellular model

Rat neocortical column

Rat brain

Human brain

1 Gigaflop   1 Teraflop   1 Petaflop   1 Exaflop

Computational speed

Deep learning neural network applied in text mining (illustrative)

y_1 y_2 y_3 y_4 y_5

D   D   D   D   D

B1   B2   B3   B4   B5

F1   F2   F3   F4   F5

C W   C W   C W   C W   C W

+ RNN (Recurrent Neural Network)

CNN
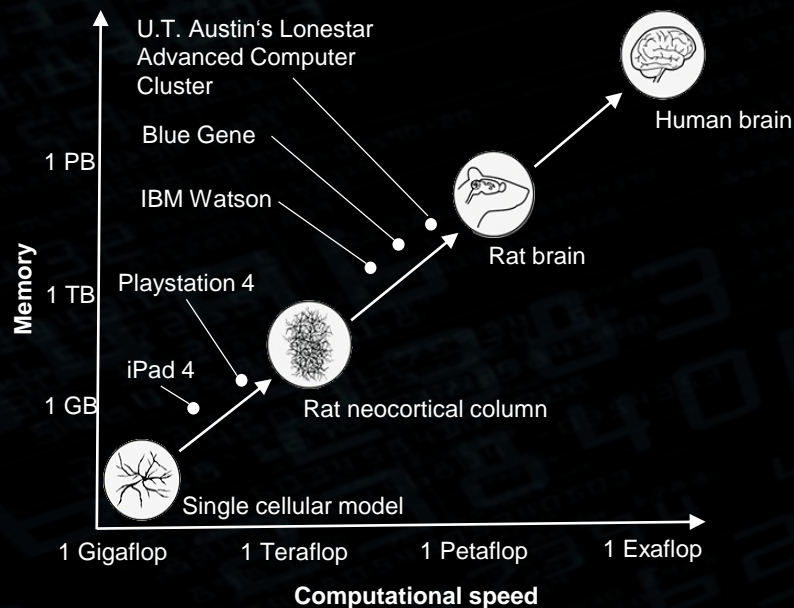
Glove / Word2vec

Text

I   saw   Einstein   in   Monaco

* Based on illustration on wired.com

TM framework captures recent AI innovations – text mining processes become more similar to human brain processes
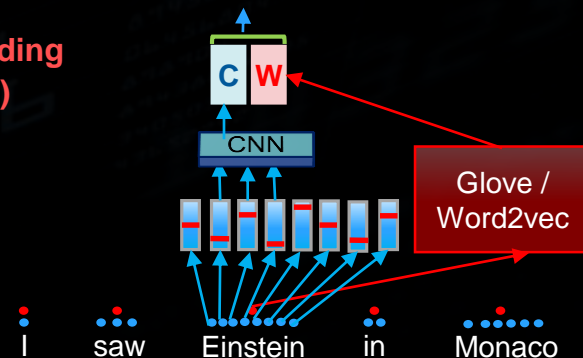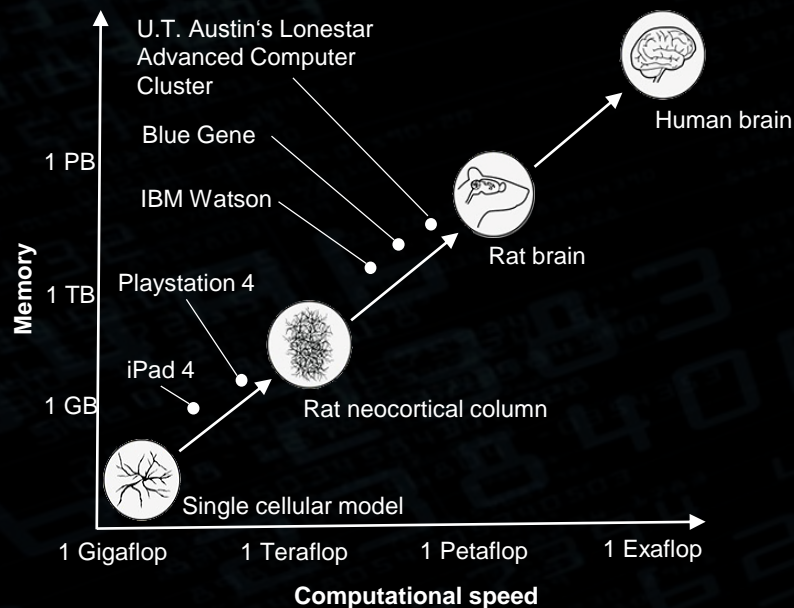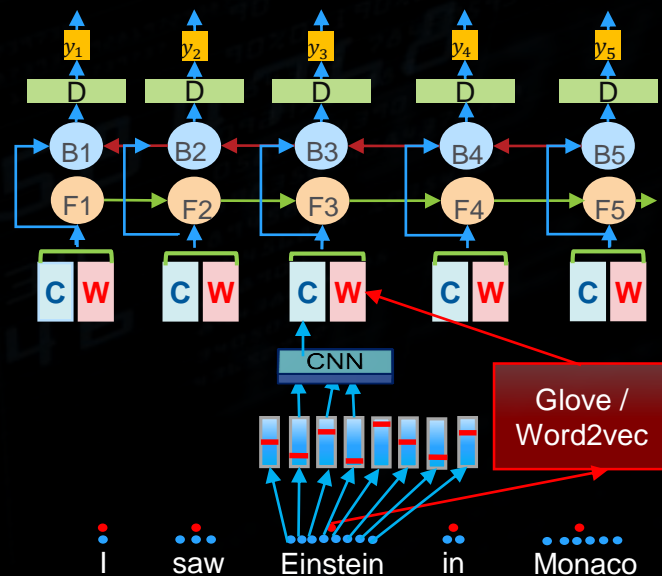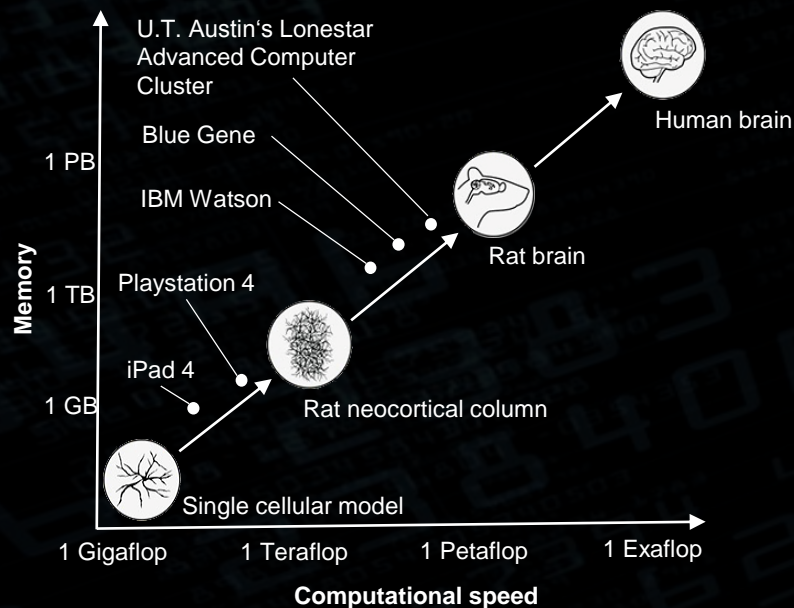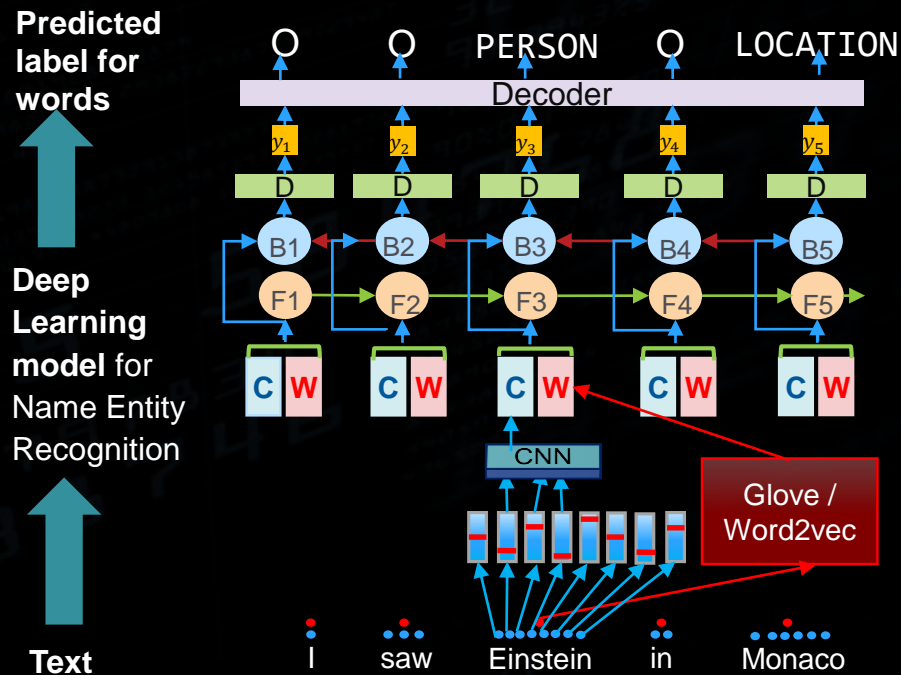
Models evolved from rule-based to those represent human brains*

Deep learning neural network applied in text mining (illustrative)

* Based on illustration on wired.com

Deep Learning moves manual work complexity to model complexity

Deep learning offers a flexible framework to approximate complex & abstract tasks by automating the feature generation process

Move manual work complexity to model complexity

	Data preparation	Feature engineering	Rules Creation	Output interpretation
Classical programming	manual			
Machine Learning	manual		automated	
Deep Learning	manual	automated		
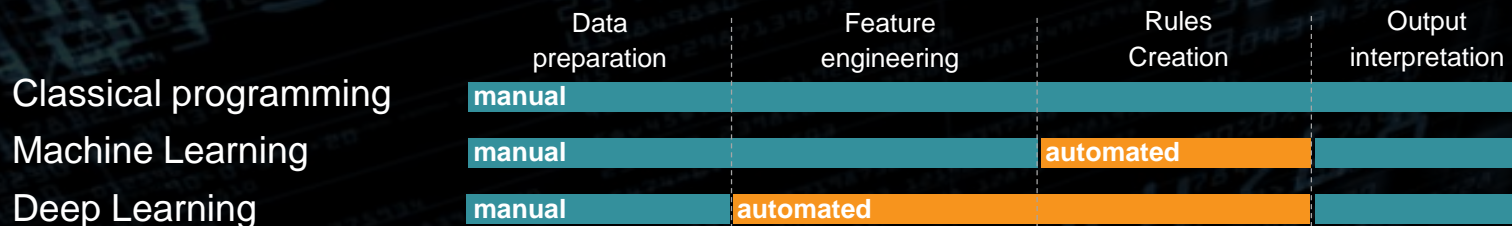
Universal approximation theorem[1]
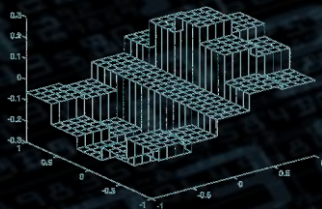
A feed-forward network with a single hidden layer containing a finite number of neurons can approximate continuous functions on compact subsets of R^n

See tasks as functions

data   function   output

        cat

image   *NN*   *class*

See complex tasks, involving structured or unstructured data as a function to be learned by the network

Hierarchical abstraction[2]

Neural network architectures typically process the data by adding complexity at each step of computation. This allows for modular re-use of models sharing similar core task

Example: image models

edges   combinations of edges   object models
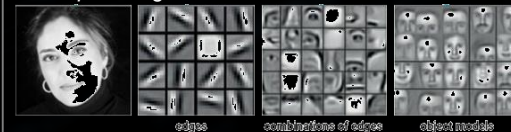
1. *Universal Approximation Using Feedforward Neural Networks: A Survey of Some Existing Methods, and Some New Results, Neural Networks, Scarselli et al., 1998*
2. *Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks, Honglak et al., ACM 2011*

3 Our approach to industrialize text mining

Scalable text mining platform: a flexible 3-tier platform, ensuring the optimal usage by internal and external business

Munich RE

Framework

| Micro-service | ⇄ | **Technology**
Using / linking existing tools in HDP, or new tools | ⇄ | **AI Logic**
Text mining modules and re-extractors | ⇄ | **Used by internal or external clients** |
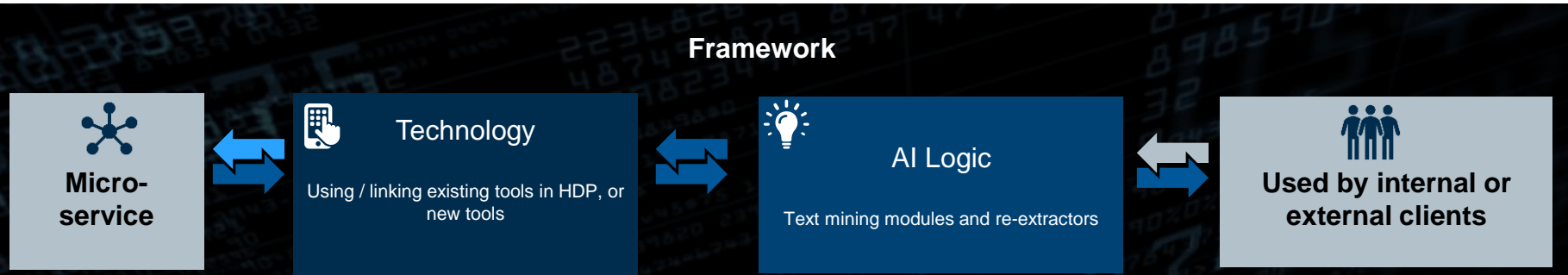
Text mining platform 3-tier setup

1 Core modules / production

2 "Managed" configure & customize

3 PoC, "self-serviced" configure & customize

- Clients can run core components as standard functionalities
- For client- specific needs, our text mining team is contracted to provide services (configuration and customization, incl. pipelines with multiple core components and data feeding)
- Clients can run PoC on platforms and evaluate performance and success of various approaches, including self-configured workflows

Annotation tool enables flexible and easy way for information capturing

Info Capturing

- Basic info
- Loss dates
- Loss location
- Risk info
- Parties involved
- Experts
- Circumstances of loss
- Legal issue
- Financials

Application example: we use text mining to detect organizational name and supply chain info

Munich RE

Text mining service to detect organizational name and supply chain info

Documents

Customized service

Visualized supply chain info and risk

Underwriters use those info for business interruption risk for companies

Company Name extractor + Site locations

Products Transportation mechanism

➡ Supply-receive relations (site to site relationship)

Risk Grading per site
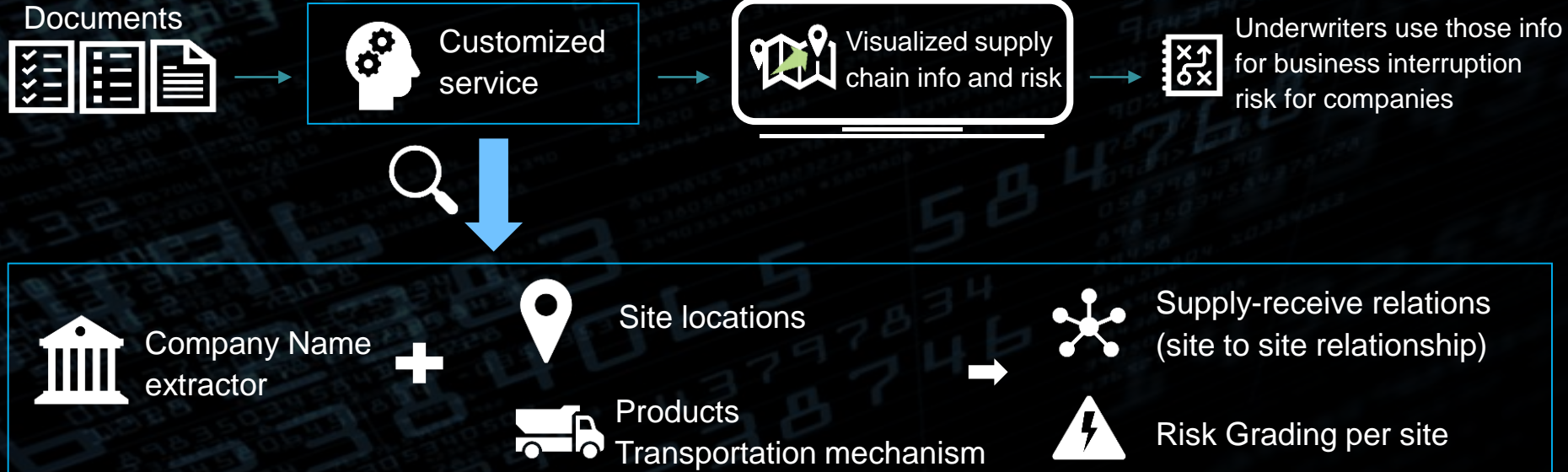
Data Lake Integration: text mining module extracts four types of info to enrich the existing documents

Info extraction

Text Extractor

txt

4 Entity Extractors
- **Company**
- **Location**
- **Person**
- **Date**

Table of Entities

Upload UI

pdf

pdf

(embedded text)

Data Lake

Keyword Index
- Company
- Location
- Company & Location

Feedback UI (check for false-positives)
- **Company**
- **Location**
- **(Person)**
- **(Date)**

pdf

View PDF

Keyword-Search UI
Example: Find all PDF files in the Data Lake having the keywords "Munich" AND "BMW"
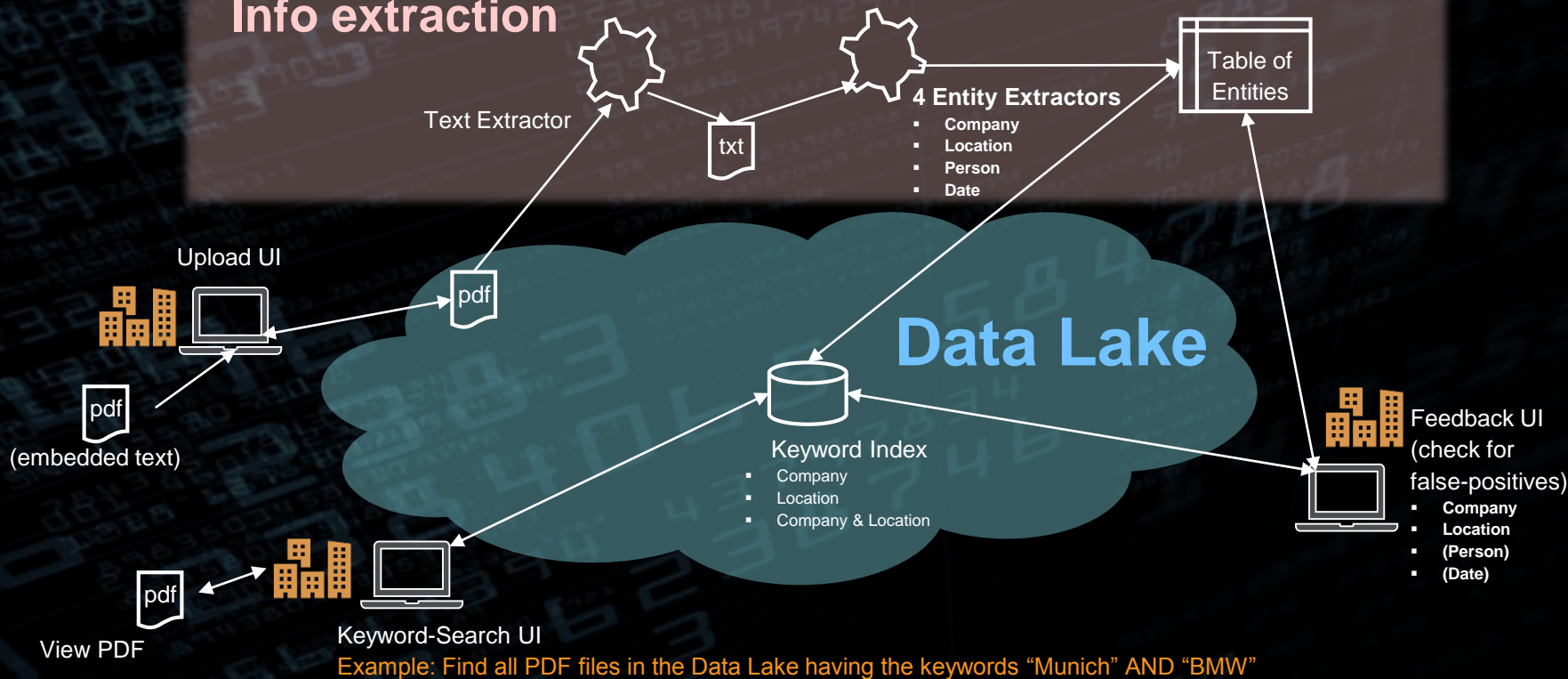
4 Long term vision

Demo: Knowledge Graph to better understand organization profile
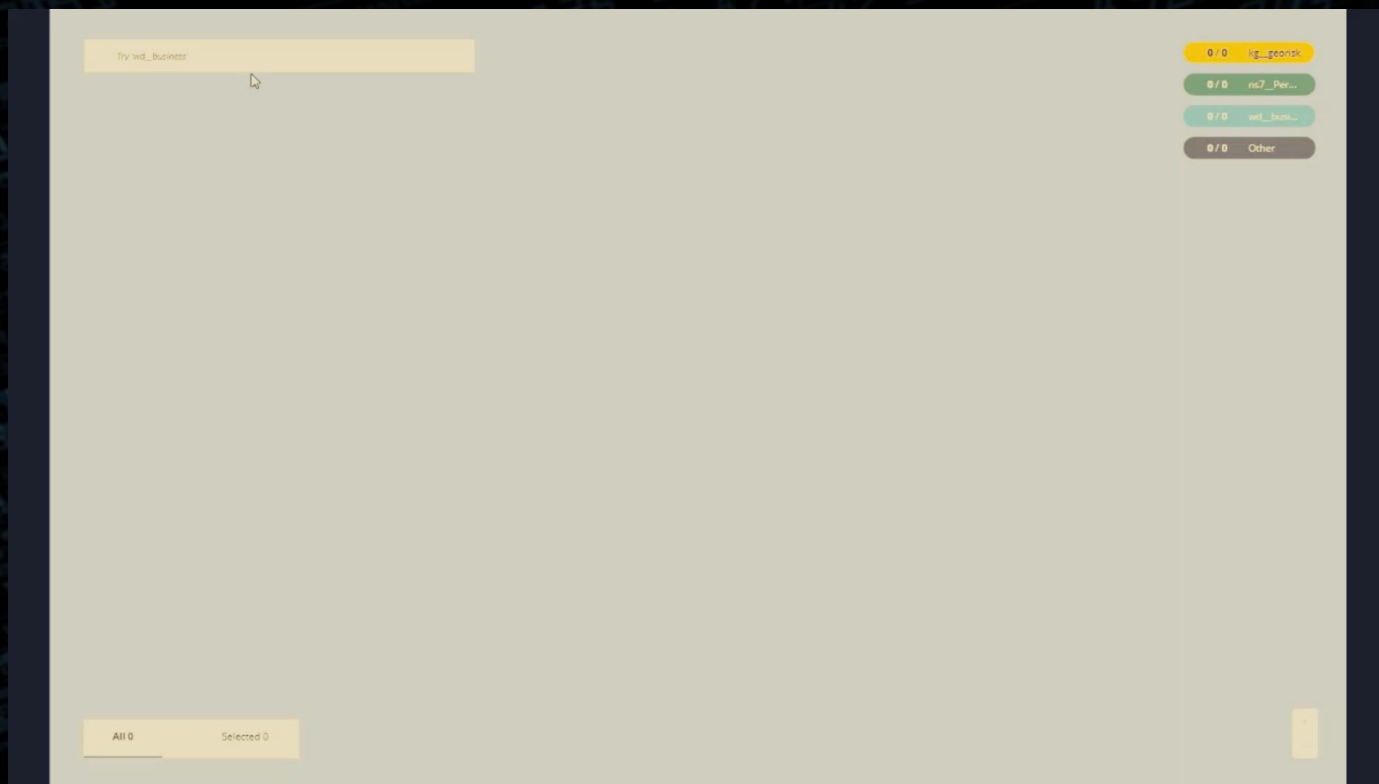
Our goal: enable access to insurance-specific and the most state-of-the-art text mining technology for both internal & external clients

Our goal

General AI

Insurance-specific AI

Microsoft
Google

Munich RE
TM Solution

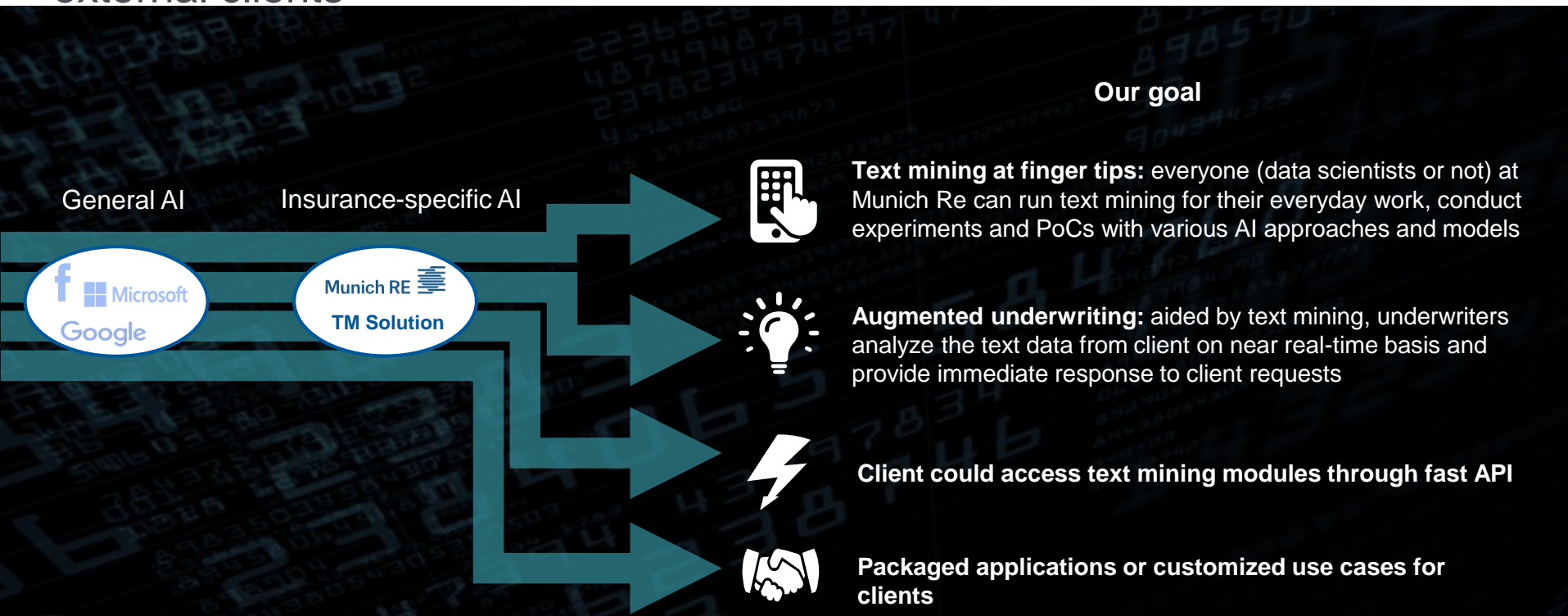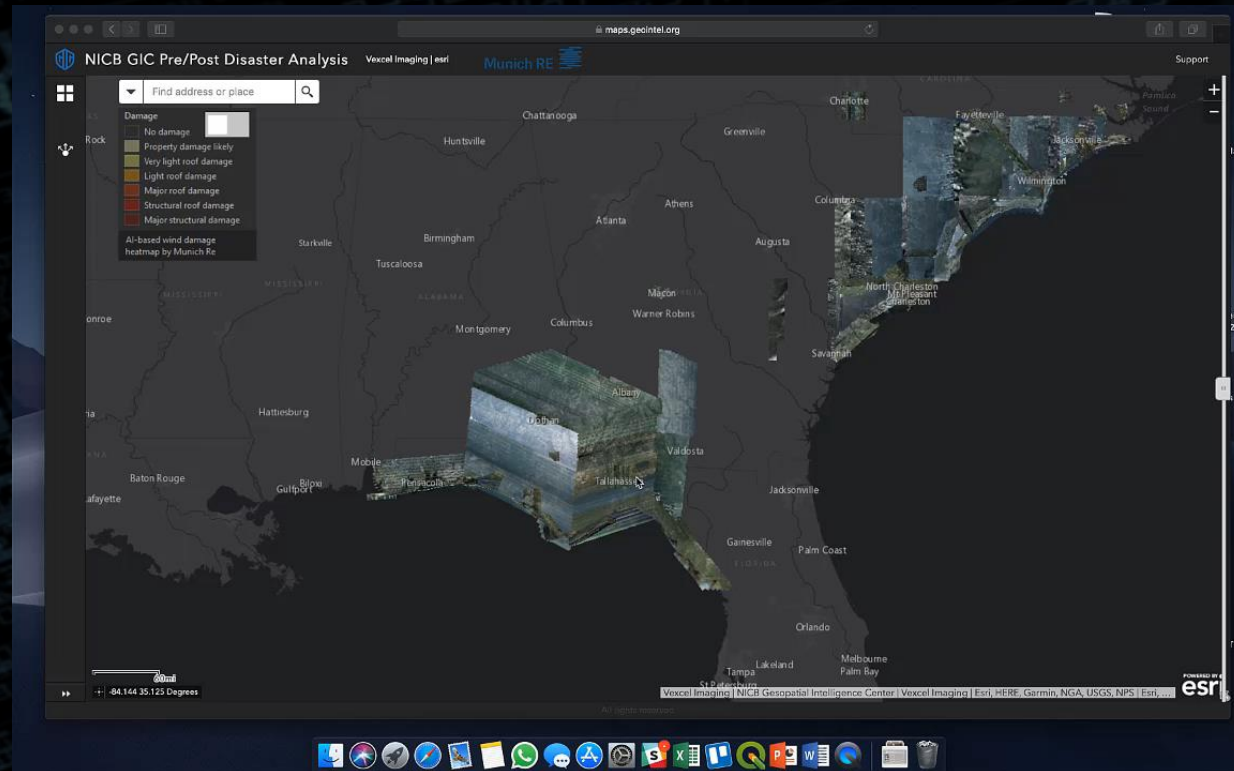**Text mining at finger tips:** everyone (data scientists or not) at Munich Re can run text mining for their everyday work, conduct experiments and PoCs with various AI approaches and models

Augmented underwriting: aided by text mining, underwriters analyze the text data from client on near real-time basis and provide immediate response to client requests

Client could access text mining modules through fast API

Packaged applications or customized use cases for clients

Demo: Deep learning approach applied for images

Question & discussion

Please feel free to contact me for further details!

Dr. Andreas Nawroth (Anawroth @ munichre.com)