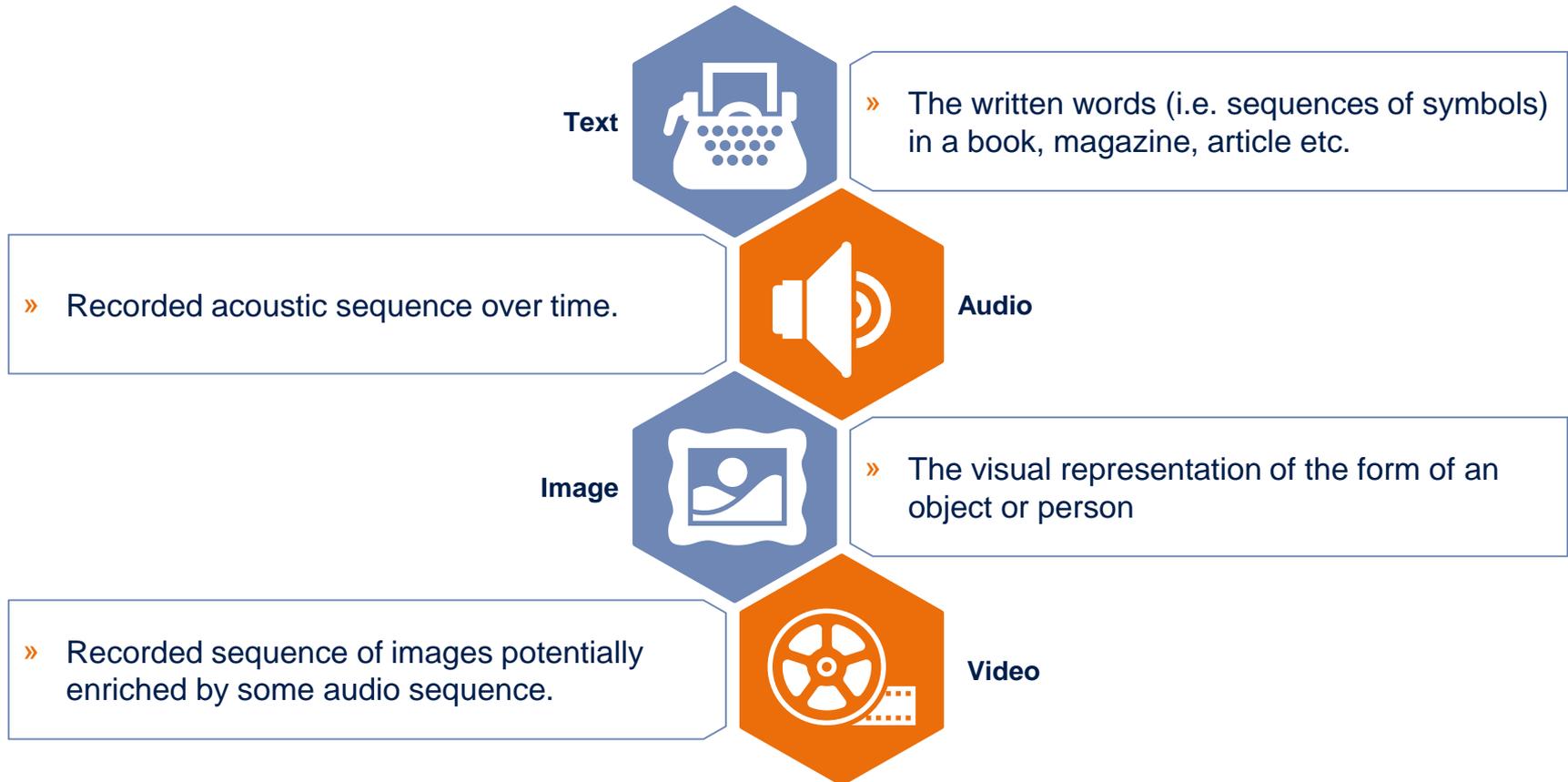


# Discovering the wilderness

Utilizing unstructured data in risk management

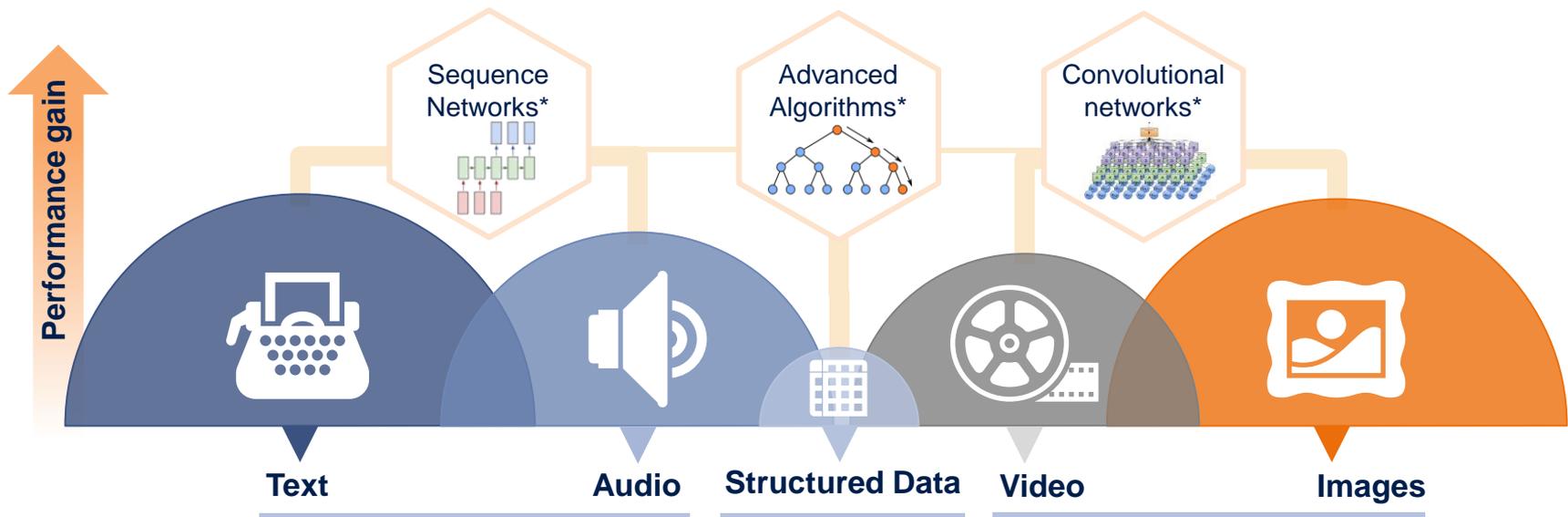
RiskMinds International, Amsterdam, 05<sup>th</sup> December, 2018

In 1998, Merrill Lynch cited a rule of thumb that somewhere around 80-90% of all potentially usable business information may originate in unstructured form



When data is unstructured, humans are (often) unable to explain their expertise on how to interpret and analyse it. Algorithms trained on examples may overcome this problem.

Machine Learning algorithms do better the more data they receive, thus tackling problems involving unstructured data is promising



- » Machine Learning has led to impressive advances such as text understanding, machine translation, image/video recognition, recommender systems or self-driven cars → **problems with unstructured data**
- » The performance enhancement gained from the application of ML algorithms to “classical” risk problems, such as credit risk prediction by **structured data** (e.g. balance sheet), is often **not worth the hassle**
- » A new risk module based on social media behaviour, however, is an ideal application

Machine Learning is not magic: What’s “*not in the data*” can also not be revealed by ML. Thus “classical” problems gain far less from ML than problems involving huge amounts unstructured data.

\* The stated algorithm classes are not exhaustive for the given problem types and are only the most prominent representatives.

# The progressing digitalization in the financial industry increased the demand for data-driven solutions, esp. for unstructured data



## Customer management

- » Acquisition and segmentation of customers
- » Improvement of customer satisfaction and loyalty
- » Strategy development for cross- and upselling



## Product management

- » Optimization of the product portfolio
- » Product engineering
- » Derivation of product recommendation and marketing strategies



## Risk/Asset management

- » Analysis, modelling and management of all kinds of (financial) risks
- » Scenario-simulations and impact analyses
- » Automated trading



## Process optimization

- » Process analysis and – optimization via prioritization and scoring
- » Agile project management via Kanban and Scrum
- » Development of data driven digital operation methods



## Network analysis

- » Identification and analysis of relationships between corporations (e.g. supplier/customer, creditor/debtor, etc.)
- » Analysis of customer networks, competitors and peer-group benchmarking



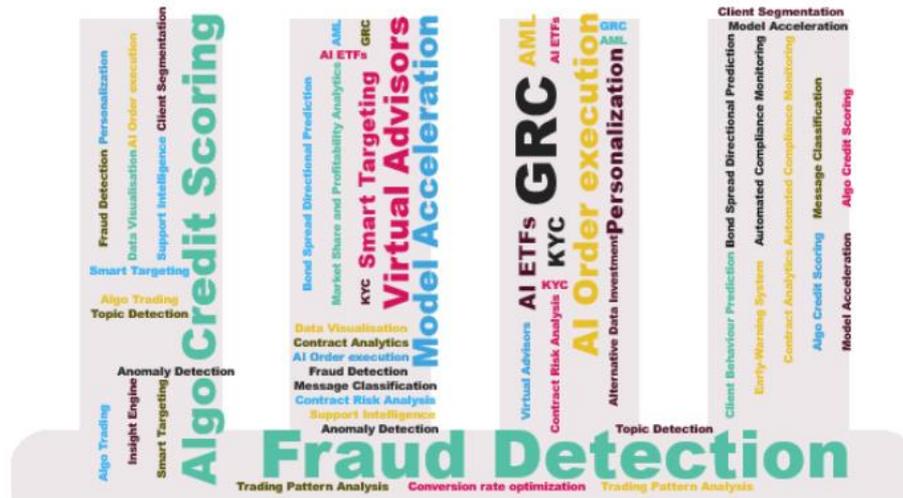
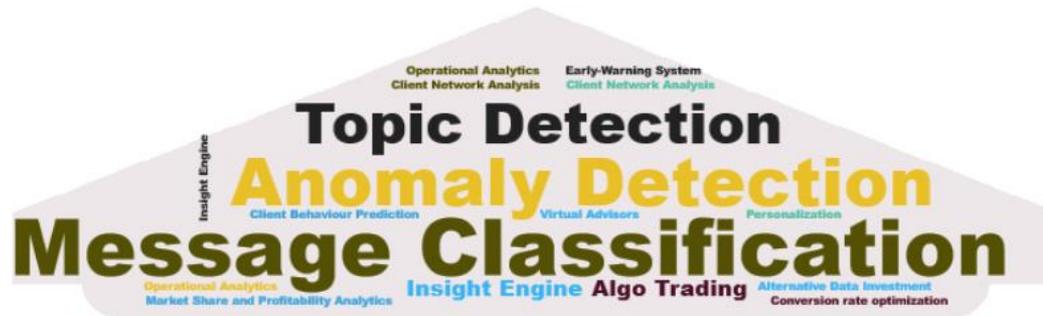
## Infrastructure

- » Analysis and optimization of the (IT-) infrastructure w.r.t Data-Science aspects
- » Competitive data collection and management
- » Efficient information retrieval

Data becomes more and more important in the everyday business life.



# Play the risk management / AI buzzword bingo



**Market Share and Profitability Analytics**  
**Bond Spread Directional Prediction**



# Machine Learning applications in Risk and Asset Management

Trading, Quantitative Investment



» Application of algorithmic trading, hedging, investment strategies and support processes for pricing

Regulatory, Compliance, Legal



» ML is used to detect abnormal financial behaviour, to cope with regulatory requirements and for the analysis of legal texts (using NLP)

Credit Scoring, Risk Management

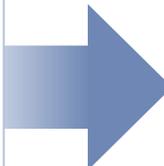
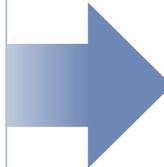


» ML is used for the assessment and monitoring of credit quality and the evaluation of triggers for risk management

Business Analytics, Business Intelligence



» Application of tools to enhance business insight using own data, e.g. by discovery of patterns, communication categorisation or visualisation



## KOALA



Universal framework for processing and analysis of **directed communication**, such as e-mails, chats, phone calls or letters





## Compliance requirements increased

- » Shortcomings in measures to prevent insider trading and market manipulation imply huge financial and reputational risks for financial institutions
- » Fines imposed by regulatory authorities on banks have been draconic, see e.g. LIBOR or Forex scandal

## Support and automatization

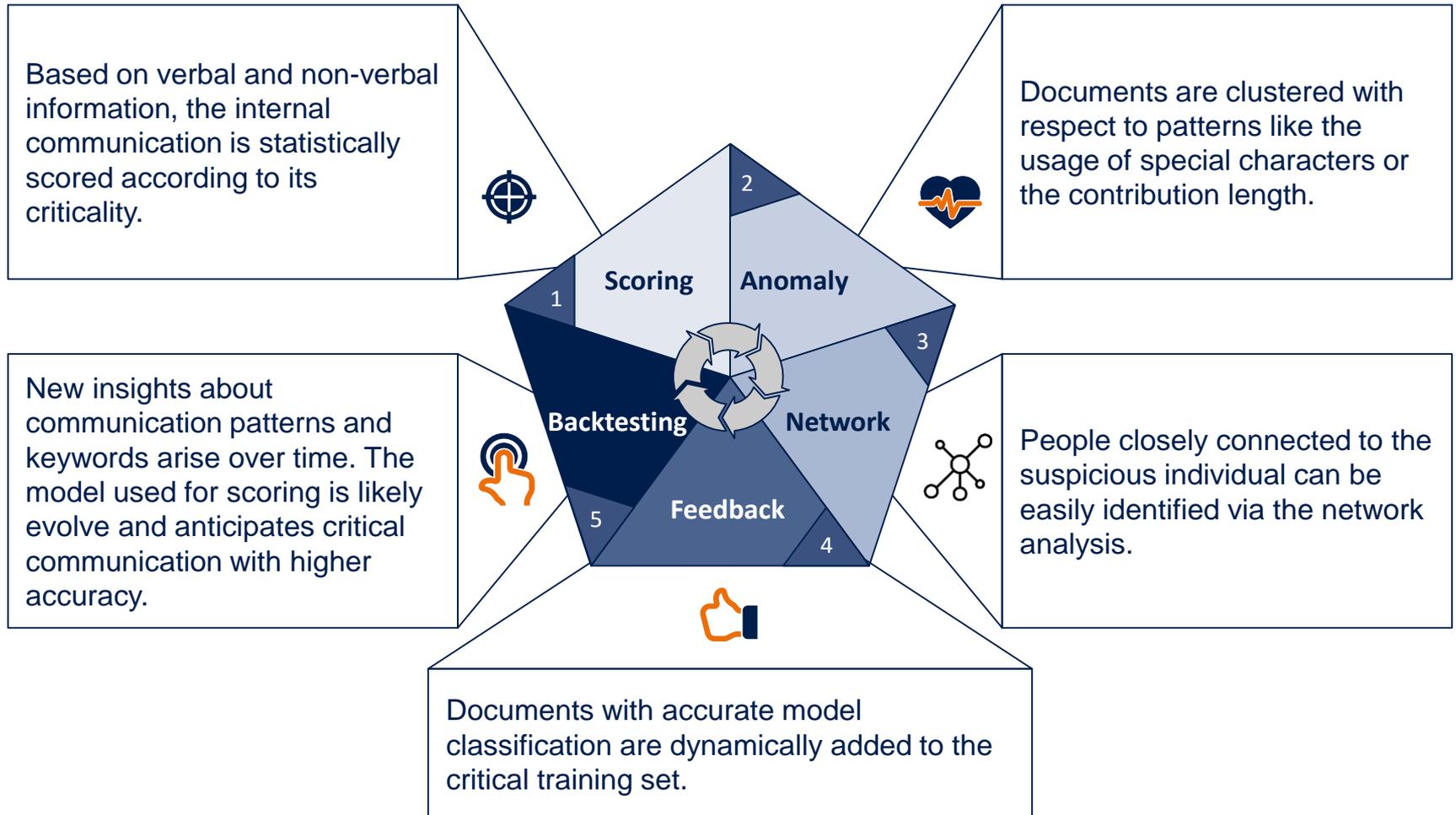
- » Based on verbal (words and word combinations) and non-verbal information (response time, number of chat participants), **communication is scored according to criticality**
- » Different information sources are jointed and analysed (e.g. trader communication. Trader positions, public news and market information)
- » Documentation requirements are met efficiently

## Internal communication in focus

- » Communication channels of traders are manifold (Email, Skype, Lync, ...), and the volume exceeds human capacities by far
- » Due to humor, sarcasm, abbreviations, different languages, spelling errors etc., internal communication is challenging to analyse



# KOALA – Our **K**ommunication **A**naLysis **A**pplication identifies critical communication and reduces compliance risks



# KOALA applied for continuous surveillance of trader communication at a leading Swiss universal bank



**Project example**

### “raw” chat logs

- » parsing raw data



### NLP processing

- » POS-tagging
- » Annotations



Chat	Feature 1	Feature 2	...
1	5	1.2	
2	4	3.5	
3	10	7.8	
...	...	...	...

### Unsupervised learning

- » Embeddings: GloVe, word2vec, LDA2vec
- » Topological data analysis
- » Deep-RNNs
- » Network Analysis



### Supervised learning

- » Naive Bayes
- » Gradient boosted trees
- » Deep-NN (CNNs)

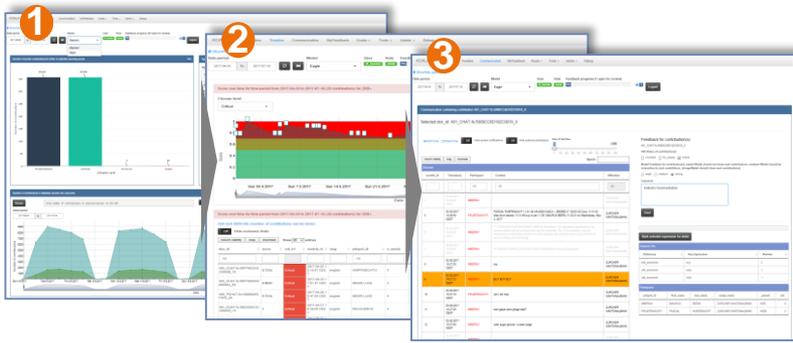


### Expert-based scoring

- » User defined keyword lists
- » Metadata analysis

### Collecting Feedback

- A** Model predictions
- B** Review by a human analyst
- C** Supervised learning



- 1** Weekly overview and statistics
- 2** Temporal evolution and criticality
- 3** Single processing of critical communication



# Machine Learning applications in Risk and Asset Management

Trading, Quantitative Investment



» Application of algorithmic trading, hedging, investment strategies and support processes for pricing

Regulatory, Compliance, Legal



» ML is used to detect abnormal financial behaviour, to cope with regulatory requirements and for the analysis of legal texts (using NLP)

Credit Scoring, Risk Management



» ML is used for the assessment and monitoring of credit quality and the evaluation of triggers for risk management

Business Analytics, Business Intelligence



» Application of tools to enhance business insight using own data, e.g. by discovery of patterns, communication categorisation or visualisation

## PANDA



Universal framework for processing and analysis of **general business communication**, such as news, contracts, financial statements or regulatory guidelines



# PANDA – Our Public Article, News and Document Analyser monitors news articles, extends trading strategies and early warning systems



## Digitalization caused a strong increase in news availability

Written information is still the most important, timely and comprehensive source for decision making in

- » Risk Management and Controlling
- » Trading and Investments
- » Research

Due to conjugation, declension, word ambiguity, negations, humour, sarcasm, abbreviations, different languages, spelling errors etc., text is not easy to analyse

## +12.000 news arrive every day

- » News from the leading news agencies (e.g. Bloomberg, Reuters, WSJ, FT) are incorporated, but local news provider may also be included
- » News articles are individually analysed for all companies in the customer's PF
- » The news history back to 2007 (in total +18m distinct news) is used as a training set

## 25% efficiency increase\*

- » The analysts' attention is directed via an ergonomic dashboard to 'important' events and news
- » Analysts may overrule the classifications, which improves future predictions
- » Processes may be automated when the methods receive a high user acceptance



\* Rough estimate, depends on the environment

# PANDA – Our Public Article, News and Document Analyser monitors news articles, extends trading strategies and early warning systems



<b>1</b>	<b>Association</b>	<ul style="list-style-type: none"><li>» Measure the strength of different relationships between entities (e.g. cooperation / competition) by analysing shared news articles</li><li>» Measure the importance of each entity for the network, simulate and measure cascades of stress</li></ul>
<b>2</b>	<b>Benchmarking</b>	<ul style="list-style-type: none"><li>» Compare companies among industries and regions</li><li>» Identify outliers resp. anomalies and early warning signals</li></ul>
<b>3</b>	<b>Clustering</b>	<ul style="list-style-type: none"><li>» Measure the sentiment / polarity etc. of news articles</li><li>» Reduce redundancy by comparing news to previously published news</li><li>» Aggregate indicators on news article level to derive a single indicator for a company over time</li></ul>
<b>4</b>	<b>Classification</b>	<ul style="list-style-type: none"><li>» Determine the most characteristic words/phrases for an entity to gauge relevant topics</li><li>» Highlight the text passages that drive the evaluation so that analysts can easily understand the implemented methods</li></ul>
<b>5</b>	<b>Time Series Analysis</b>	<ul style="list-style-type: none"><li>» Identify pattern and periods with abnormal/significant media attention</li><li>» Filter on specific news providers and time periods</li><li>» Benchmark media attention for a company against peer competitors to identify company-specific events</li></ul>
<b>6</b>	<b>Highlighting</b>	<ul style="list-style-type: none"><li>» Guide the analyst towards relevant news items respectively passages</li><li>» Reduce redundancy among documents via word-tree structures</li><li>» Make the model classification transparent via colouring</li></ul>



# PANDA applied for text mining/information extraction and process automation at a large Austrian banking group



**Project example**

## Data clustering & extraction

- » Identify common formats and structures guarantee applications
- » Extract relevant information (e.g. customer name and address, amount, currency, type of guarantee and expiry, beneficiary, ...)



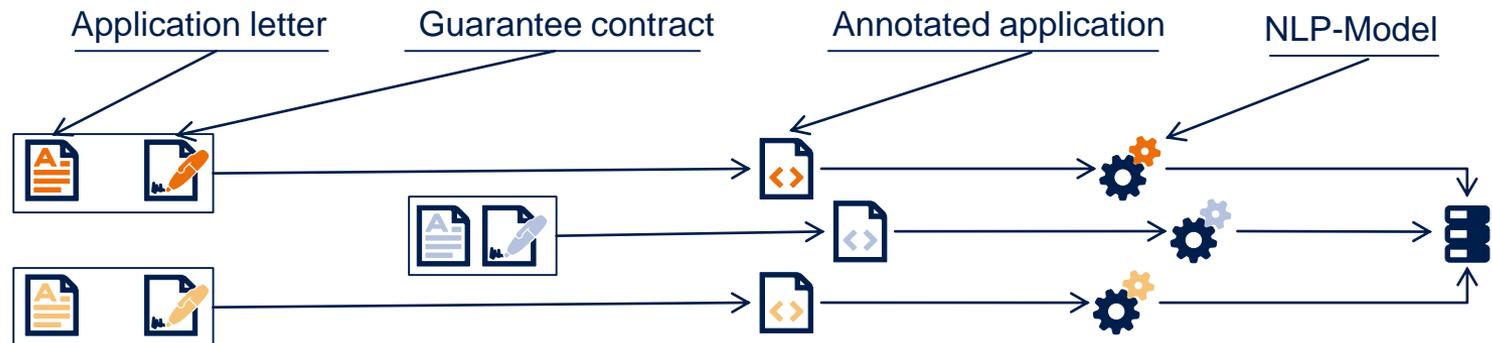
## Data validation & classification

- A** Model predictions
- B** Review by a human analyst
- C** Automated model enhancement



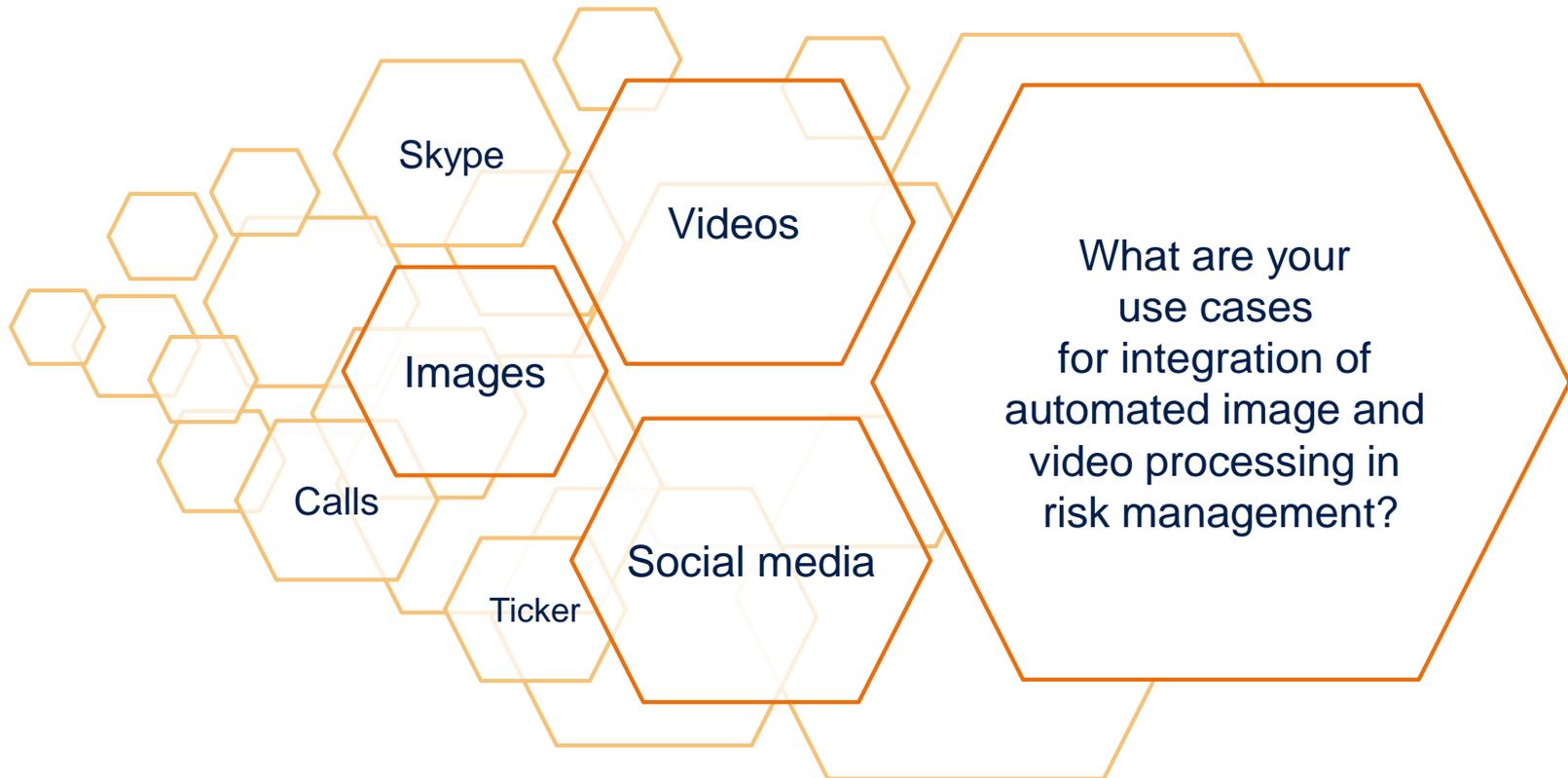
## Process automation

- » Scoring of guarantee application process according to its criticality / complexity
- » Process automation in Blue Prism (approval, request additional information, rejection)
- » Communication with the front office is transferred by a chatbot



Automatic processing of text with Machine Learning techniques is already present in risk management. What about other unstructured data?

---



d-fine