

A night-time photograph of the Tower Bridge in London, illuminated with warm lights against a dark blue sky. The bridge's two towers and suspension cables are clearly visible. In the foreground, a dark metal railing with a diamond-patterned mesh runs across the frame, and a paved walkway is visible on the left.

# GRC

SUMMIT 2022

LONDON, NOV 8-9

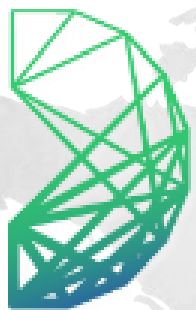
Hosted by MetricStream

## How **Artificial Intelligence** supports GRC: Text analysis use cases applied to Operational Risks

Michele Filannino

Principal Data Scientist @ Prometeia spa





# prometeia

**1000+**  
PEOPLE

**500+**  
CLIENTS

## BRANCHES

- Istanbul
- Cairo
- London
- Moscow
- Zurich

## OFFICES

- Bologna
- Milan
- Rome

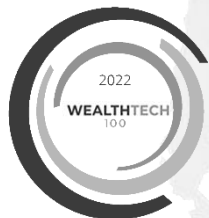
## PARTNERSHIPS

- Amman
- Athens
- Dubai
- Kuwait City
- Lagos
- Lomè
- Madrid
- Paris
- Stockholm
- Tunis
- Vienna



**RiskTech  
100 2022**

Prometeia  
Balance Sheet  
Risk Management



**XCELENT  
Technology 2020**



**REFINITIV**  
Top 25 companies  
for M&A deals in Italy



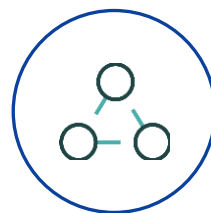
# GRC transformation: towards a new approach

**In a context of customers' rising expectations, competitive pressure, regulatory constraints, emerging risks and new business models, GRC functions need to accelerate their digitalization efforts, moving towards a new approach that should be:**



## PROACTIVE

Evolving from a traditional and reactive approach to one that is proactive, improving the organization's ability to avoid or manage both existing and emerging risks, aiming for an effective cost of risk reduction



## INTERCONNECTED

Understanding and analysing interrelationships and dependencies among various types of risks, collaborating with every part of the organization and with a strong connection with the first line of defence and frontline employees



## DATA AND TECH-DRIVEN

Reimagining control functions with a focus on data and analytics, using large amounts of data produced by digital processes (digital traces), through high-performance and scalable machine learning solutions



## VALUE-ORIENTED

Supporting, according to their role, a risk-controlled and long-term growth of corporate value, providing insights to business, sustaining executives' decisions and unlocking a range of organizational benefits



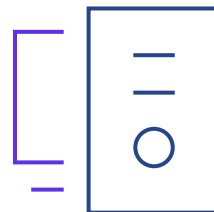
# What GRC needs to complete the transformation

Despite the investments in the area of risk mitigation and control, in our vision, the transformation process is still ongoing and requires the presence of the following enablers to become effective



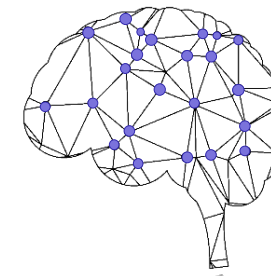
## HARD SPECIALIZED SKILLS

The continue growth in complexity of business models and the spreading of emerging risks require a plurality of differentiated specialized skills (data science, cyber risk, environmental sciences...)



## PLATFORMS AND SOLUTIONS

The development and adoption of suitable application solutions, for managing records, deadlines, workflows, documents, integrated with advanced analytics, allow the effective risk control and losses prevention



## ARTIFICIAL INTELLIGENCE

The analysis of structured and unstructured data, using artificial intelligence tools (machine learning, natural language processing...), provides a reliable risks quantification supporting business decisions, according to a data-driven approach

# structured **vs** unstructured data



relational  
DBs



records

~40%

already available in a structure suitable for analysis



CVs



reports



emails



web  
pages



PDFs



videos



social  
posts



images

~60%

not ready to be analysed, typically stored in their original formats

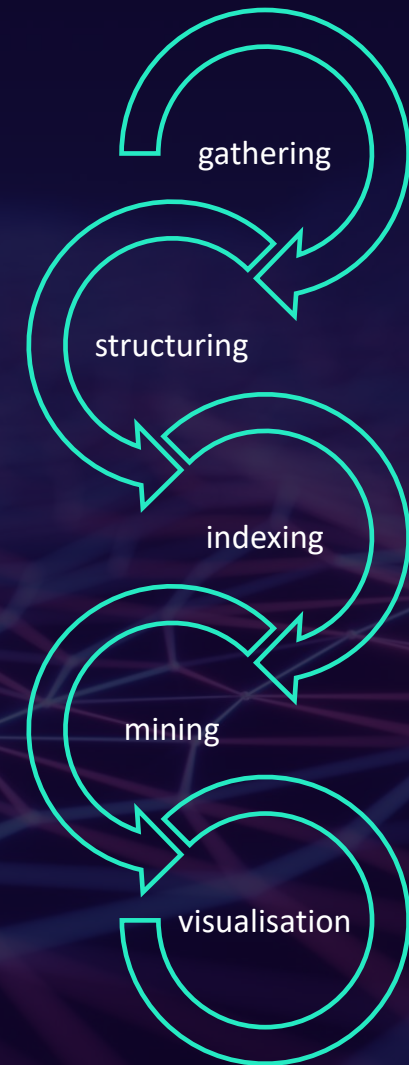
Source: <http://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/>

# the computing journey: from text to **insights**

A series of task performed to enable an efficient extraction of insights from textual data.

From a functional point of view, such activities are:

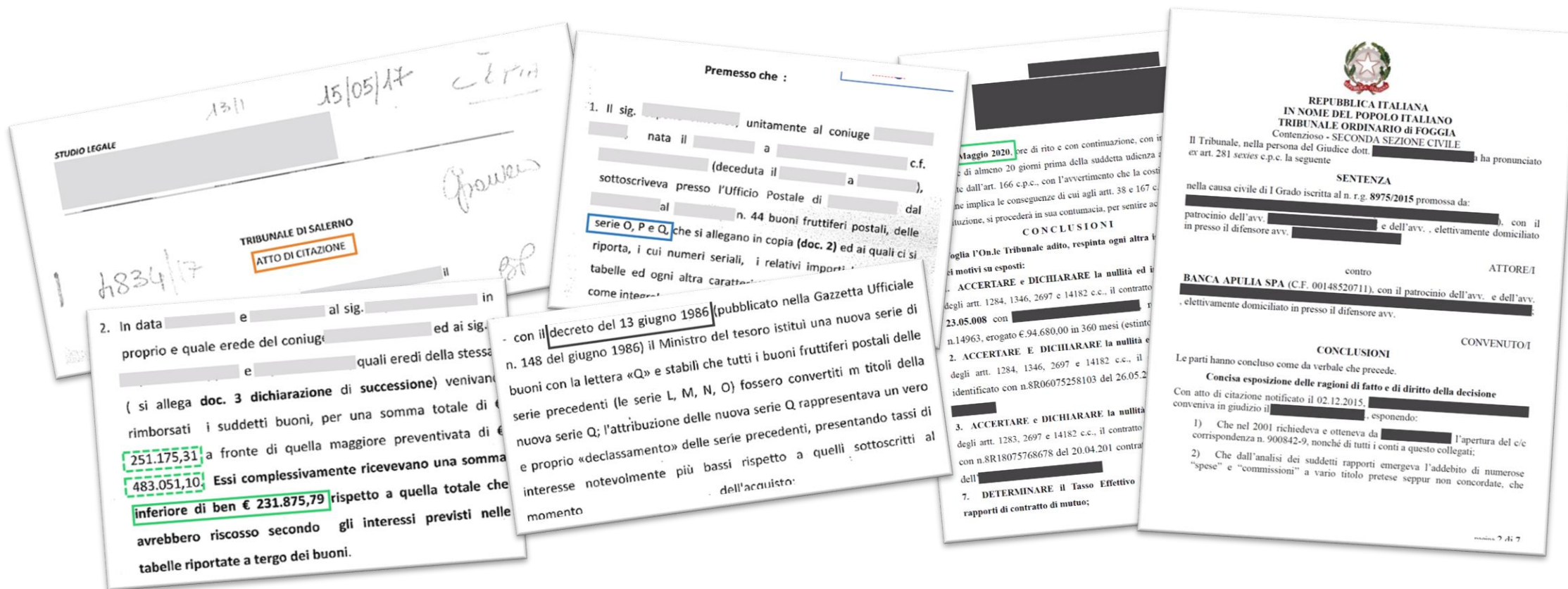
- **gathering**: it's about selecting and assembling the right textual resources;
- **structuring**: preparing heterogeneous textual sources for a more suitable machine-understandable representation;
- **indexing**: storing the representations so that such texts are easily accessible and searchable;
- **mining**: performing analytics, factual inference and/or information extraction;
- **visualisation**: once the information is structured, often insights arise by visualising data in context.





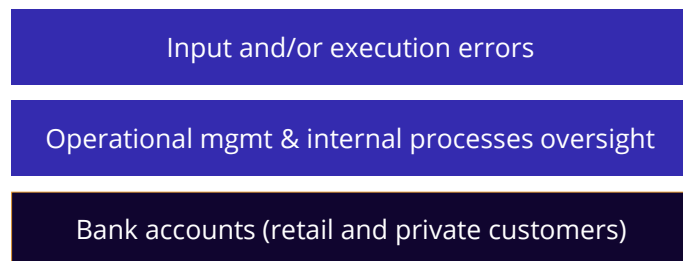
# advanced layout analysis and OCR

Text often needs to be extracted from documents, forms, attachments. This is the case for certain types of claims, legal documents, contracts, policies, regulations, process & control descriptions.



# information extraction

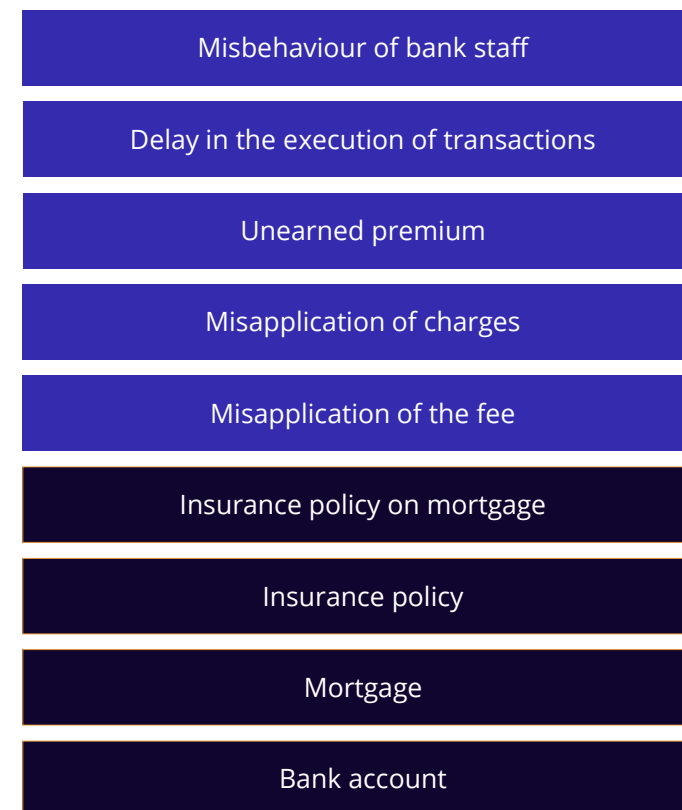
## PRE-EXISTING CLASSIFICATION



■ Root causes ■ Products

Dispute over delays attributable to branch in claiming untaken premium of N4 Mortgage insurance policy and claiming refund of current account fee December 2017. Mr. Ross complains about the branch's conduct in the context of his request for repayment of the untaken premium of the N4 Mutuo policy, following the repayment of the underlying loan. The customer claims that the branch behaved unprofessionally and initiated the file late. Since the account cannot be extinguished at this time pending the crediting of the refund, the customer requests a refund of the December fee.

## FINE-GRAINED TAGGER





# linguistic trend analysis

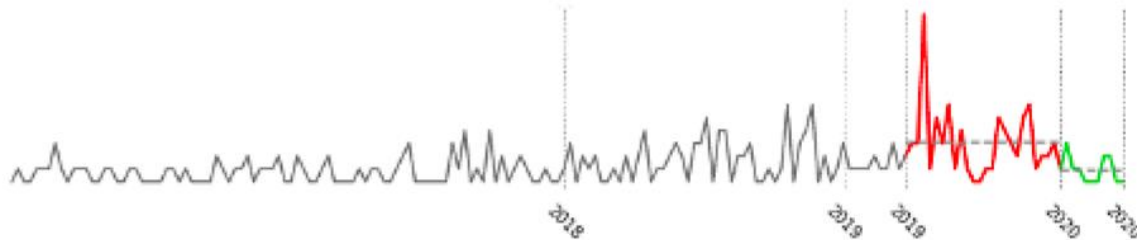


Root cause

## DISPUTED BANK TRANSFER

The number of events from April 2020 **diminished** on average of 2.1 events per week as opposed to the previous time period (October 2019 - April 2020).

Sensibility: 1 ————— 10



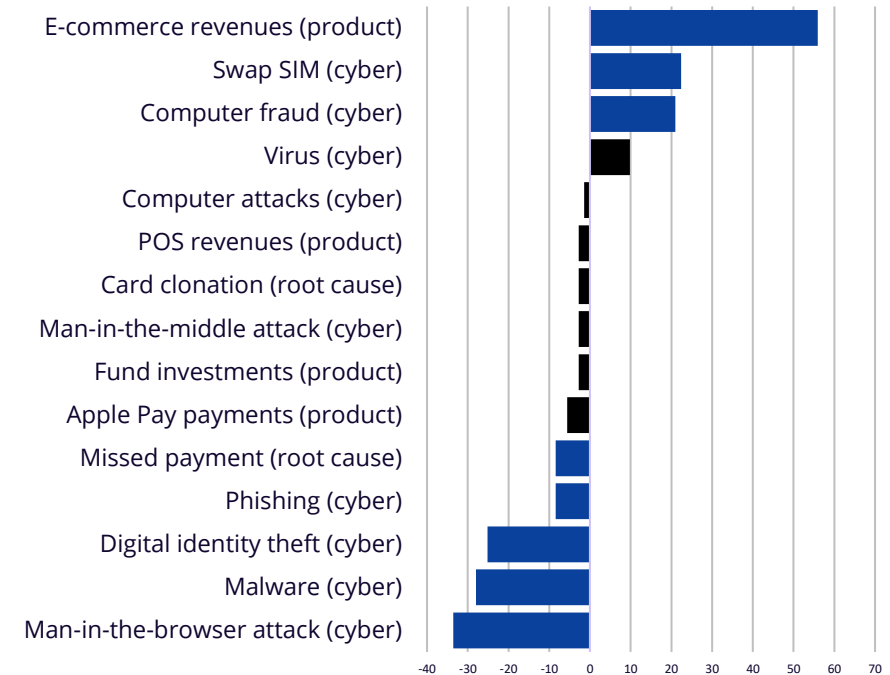
Trend

**DOWNWARD**

Temporal window  
Jan. 2017 – Jun. 2020

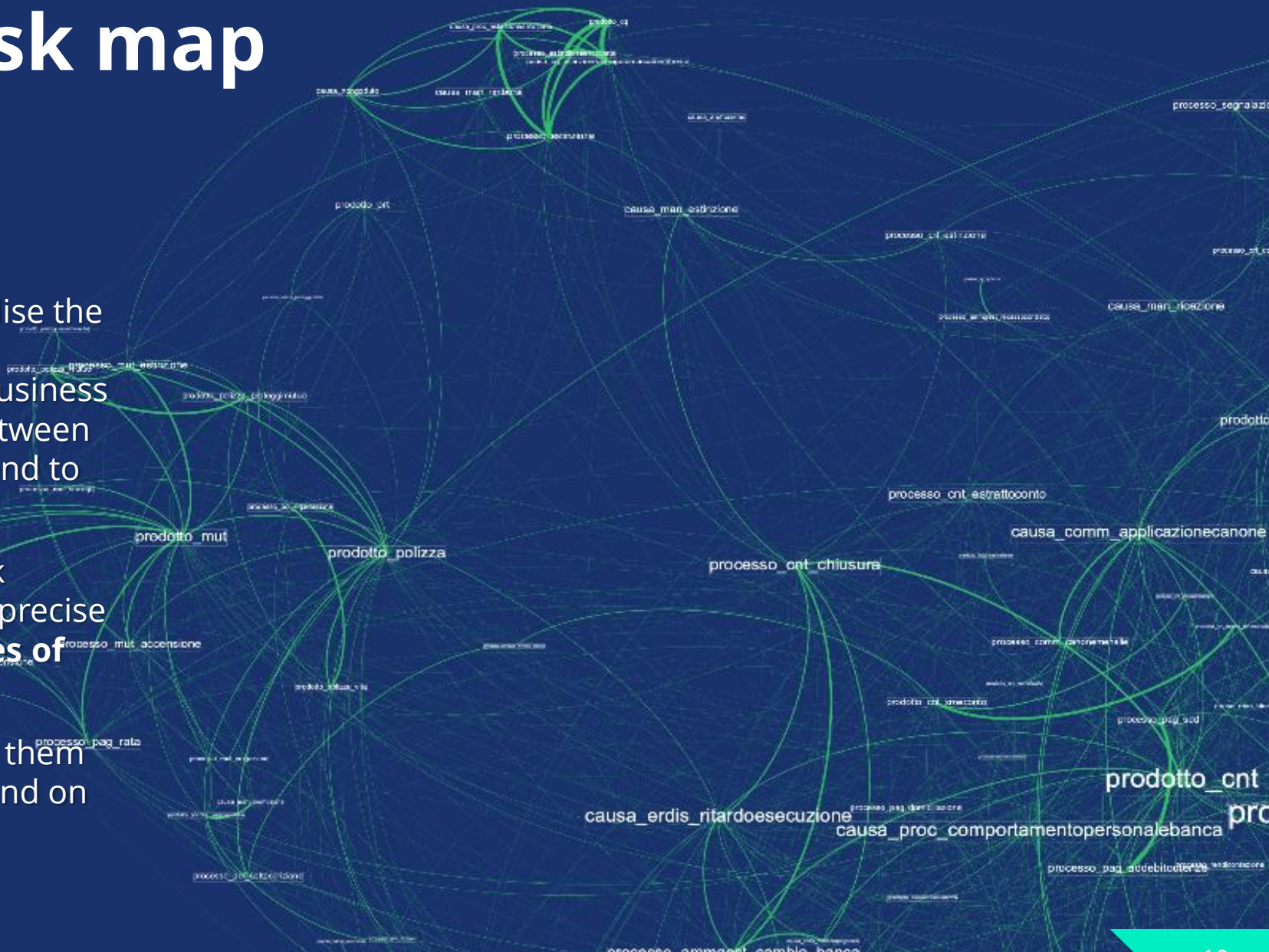


## ANALYSIS CO-TRENDS



# operational risk map

- The tagging can be used to visualise the overall map of generated business concepts. Each node represents a business concept whereas the connection between them is as strong as the concepts tend to occur together in losses notes
- The map enables operational risk managers to get an immediate and precise characterization of the **main sources of risks**
- Exploring such map also enables them to identify unexpected losses and land on peculiar cases



# data quality

In connection with **lease agreement** No. 123 it is requested to: - to ascertain and declare the nullity of the contractual clauses relating to the interest rate, as they provide **usurious conditions** [...] This, by virtue [...] and of the **compounding interest** effect resulting from the adoption of a financial plan with so-called French amortization; [...] in the alternative, declare the interest clause null and void because it violates Law 287/1990 and Article 117 TUB due to the adoption of Euribor (a value alleged to be the object of a cartel among banking institutions) as a variable parameter in violation of **antitrust regulations**, with application in place of BOT rates [...]

## TAGS

compound interest

usury

antitrust

## AUTOMATIC PREDICTION

### PREDICTED EVENT TYPE

ET 04.01.02 Usury

### GIVEN EVENT TYPE

ET 04.01.02 Usury

## SUGGESTED ACTION



The **employee**, a manager of the \*\*\* branch, received a **report contesting violations** of the combined provisions of Articles 2 , paragraph 152 of Legislative Decree No. 262/2006 and 5 of the April 28, 2016 DM of the Ministry of Economy and Finance. Specifically, this concerns the late **reporting to the MEF** of suspected **counterfeit banknotes**

MEF communication

employee's  
infringement

fake banknotes

### PREDICTED EVENT TYPE

ET 07.02.01 Supervisory monitoring & reporting

### GIVEN EVENT TYPE

ET 07.01.10 Input and/or execution errors



CONFIDENCE  
85.9%



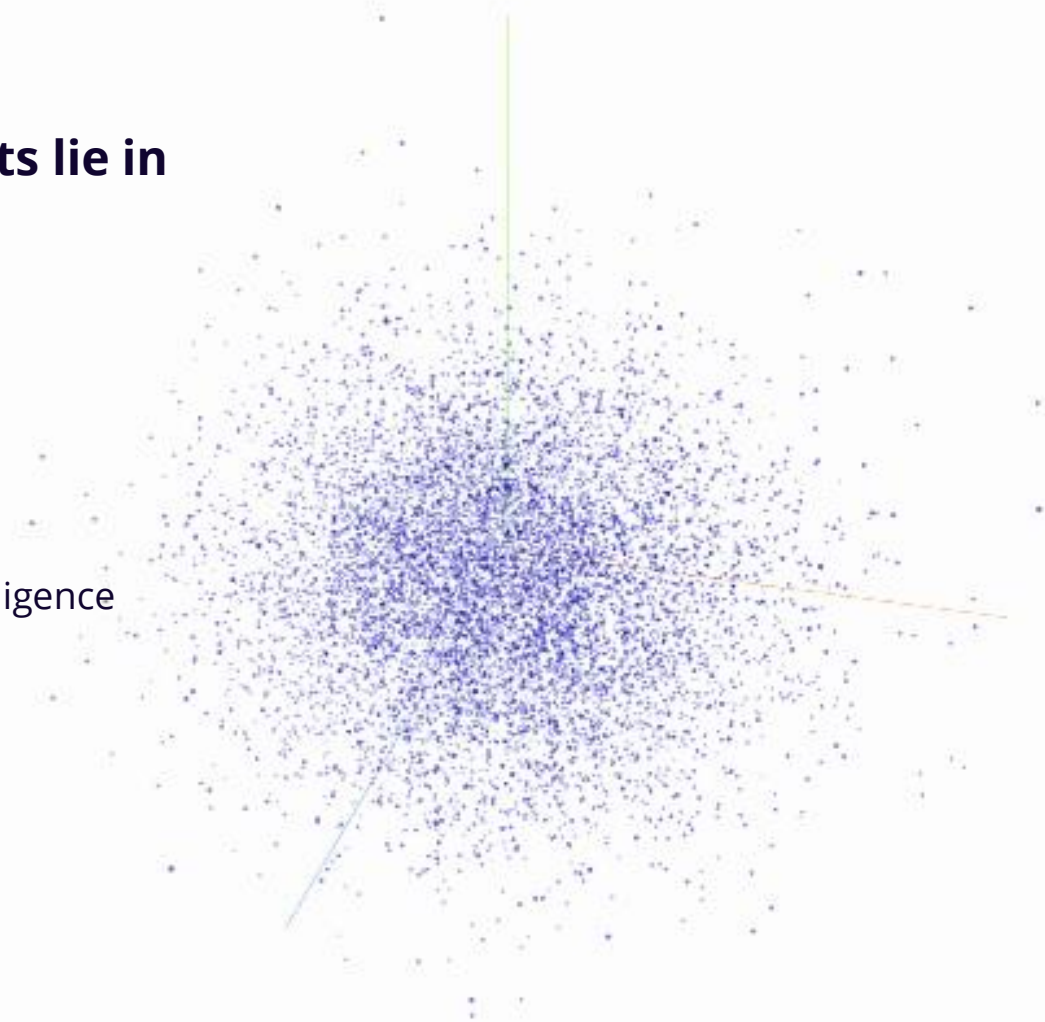


under the hood

# large language models

complex algebraic spaces where concepts lie in clusters of similar meaning

- **ad-hoc** language models
- tag-based representation
  - accelerated and improved via artificial intelligence
- tailored **information design**
- advanced statistics expertise
- expert-in-the-loop approach





# text analysis applications are already mature!

## TEXT CATEGORIZATION



**Input:** corpora, categories  
**Output:** category per document

## TOPIC MODELLING



**Input:** corpora  
**Output:** topics per document

## QUESTION ANSWERING



**Input:** corpora, questions  
**Output:** pertinent answer

## DOCUMENT SUMMARIZATION



**Input:** document  
**Output:** shorter document

## MACHINE TRANSLATION



**Input:** sentence in a language  
**Output:** sentence in a different language

## NAMED ENTITY RECOGNITION



**Input:** document  
**Output:** categorized entities

## CHATBOT



**Input:** corpora, requests  
**Output:** satisfactory answers (per request)

## SENTIMENT ANALYSIS



**Input:** corpora  
**Output:** set of authors' feelings

## INFORMATION RETRIEVAL



**Input:** corpora, query  
**Output:** ranked list of relevant documents



A night-time photograph of the Tower Bridge in London, illuminated with warm lights against a dark blue sky. The bridge's iconic towers and suspension cables are clearly visible.

# GRC

## SUMMIT 2022

LONDON, NOV 8-9

Hosted by **MetricStream**

**Michele Filannino**

Principal Data Scientist @ Prometeia spa

[michele.filannino@prometeia.com](mailto:michele.filannino@prometeia.com)

