



Efficient Explainable Learning on Knowledge Graphs

AI on Demand Technical Contributors Board Meeting - 23/02/24



Funded by
the European Union

Partners: Universitaet Paderborn
Universiteit van Amsterdam
National Centre for Scientific Research “Demokritos”
European Union Satellite Centre
DATEV EG
Weblyzard Technology GmbH

Type of Action: HORIZON Research and Innovation Actions

Topic: HORIZON-CL4-2021-HUMAN-01-01

Project duration: October 1 2022 - September 30 2025 (36 months)

Website: <http://enexa.eu>

Github: <https://github.com/EnexaProject>

Linkedin: <https://www.linkedin.com/company/enexa-eu-project/>

Twitter: https://twitter.com/enexa_eu

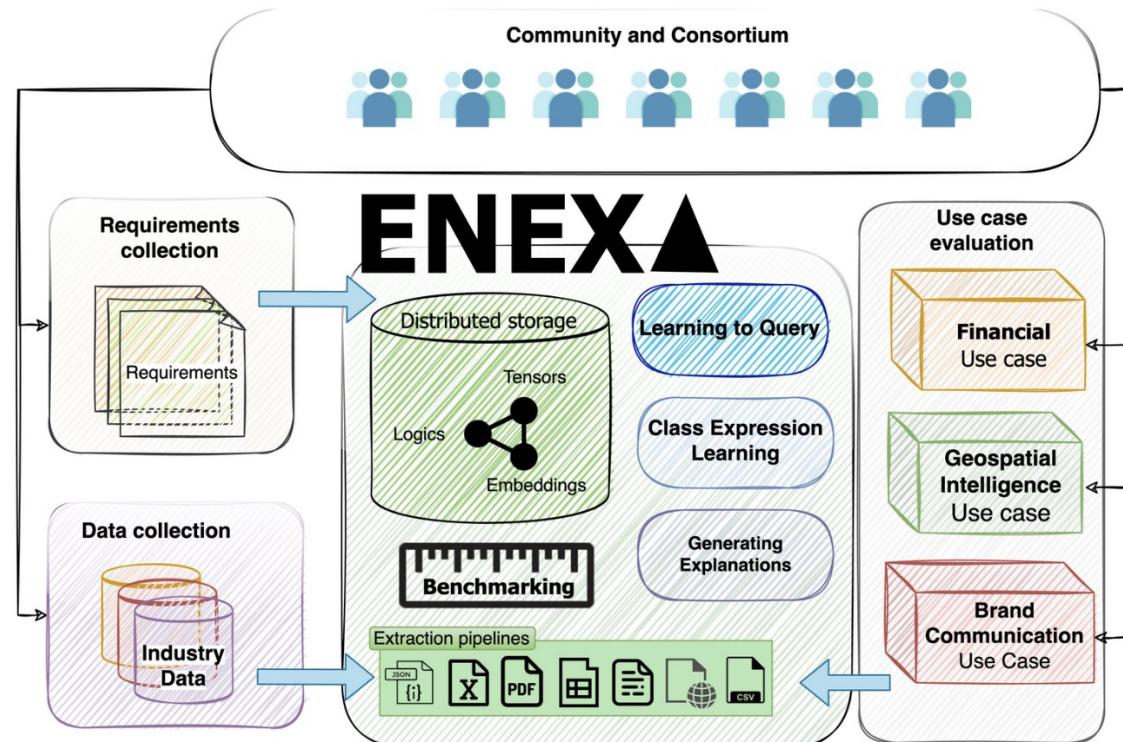


Motivation and Project Ambition

- Real-world Knowledge Graphs (KG) are being used in a variety of domains
 - ▽ Web search, question answering, personal assistants, XAI across most sectors including finance, Industry 4.0, personalised medicine, legislation and more
- Current AI and ML algorithms for KGs:
 - ▽ Fail to scale for very large sizes (billions of edges)
 - ▽ Are not useful in cases of formal inconsistency
 - ▽ Fail to exploit the semantics modeled into the KGs
 - ▽ Rely on a one-shot explanation paradigm—if they are at all explainable.
- Main ENEXA Objectives:
 - Explainable and Hybrid ML for real-world KGs with a focus on human-centered explanations
 - Representation trichotomy: Tensors, Embeddings, Formal Logics
 - Co-construction of Explanations: Hybrid ML for human-centered explanations

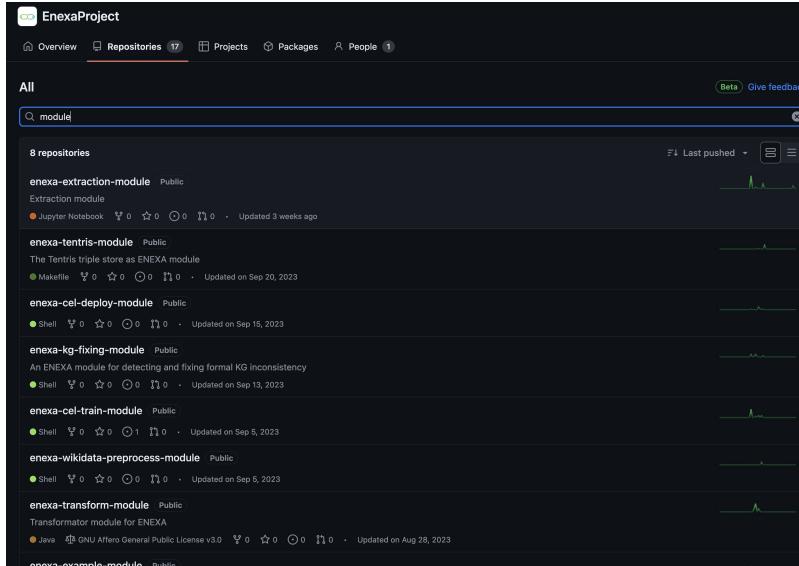
Overall Architecture of ENEXA

- Modular
- Reproducible
- Explainable



Results so far

- Sourcecode <https://github.com/EnexaProject>



- Publications <https://enexa.eu/publications/>

- Online Event Recognition over Noisy Data Streams.** P. Mantenoglou, A. Artikis, and G. Paliouras. *International Journal of Approximate Reasoning*, Elsevier.
- Logics with probabilistic team semantics and the Boolean negation.** Miika Hannula, Minna Hirvonen, Juha Kontinen, Yasir Mahmood, Arne Meier, Jonni Virtema. *JELIA 2023*
- Learning Permutation-Invariant Embeddings for Description Logic Concepts.** Caglar Demir, Axel-Cyrille Ngonga Ngomo. *IDA 2023*
- Native Execution of GraphQL Queries over RDF Graphs Using Multi-way Joins.** Nikolaos Karalis, Alexander Bigerl, Axel-Cyrille Ngonga Ngomo. *SEMANTECS 2023*
- TemporalFC: A Temporal Fact Checking approach over Knowledge Graphs.** Umair Qudus, Michael Röder, Sabrina Kirrane, Axel-Cyrille Ngonga Ngomo. *ISWC 2023*
- Quantitative Reasoning and Structural Complexity for Claim-Centric Argumentation.** Johannes K. Fichte, Markus Hecher, Arne Meier, and Yasir Mahmood. *IJCAI 2023*
- Neuro-Symbolic Class Expression Learning.** Caglar Demir, Axel-Cyrille Ngonga Ngomo. *IJCAI 2023*
- Clifford Embeddings – A Generalized Approach for Embedding in Normed Algebras.** Caglar Demir, Axel-Cyrille Ngonga Ngomo. *ECML PKDD 2023*
- Parameterized Complexity of Propositional Inclusion and Independence Logic.** Yasir Mahmood and Jonni Virtema. *WoLLIC 2023*
- Neural Class Expression Synthesis in ALCHIQ(D).** N'Dah Jean Kouagou, Stefan Heindorf, Caglar Demir, Axel-Cyrille Ngonga Ngomo. *ECML PKDD 2023*
- Explainable Integration of Knowledge Graphs using Large Language Models.** Abdullah Fathi Ahmed, Asep Fajar Firmansyah, Mohamed Ahmed Sherif, Diego Moussallem, and Axel-Cyrille Ngonga Ngomo. *NLDB 2023*
- Neural Class Expression Synthesis.** N'Dah Jean Kouagou, Stefan Heindorf, Caglar Demir, Axel-Cyrille Ngonga Ngomo. *NeSys 2023*
- Data journeys: Explaining AI workflows through abstraction.** Enrico Daga and Paul Groth. *Semantic Web Journal 2023*



Use cases

23rd February 2023



Funded by
the European Union



Use Case 1
Business software solutions
DATEV eG



Funded by
the European Union

DATEV eG

- An IT community for tax advisors, auditors, lawyers and their clients
- One of Europe's largest IT service providers *

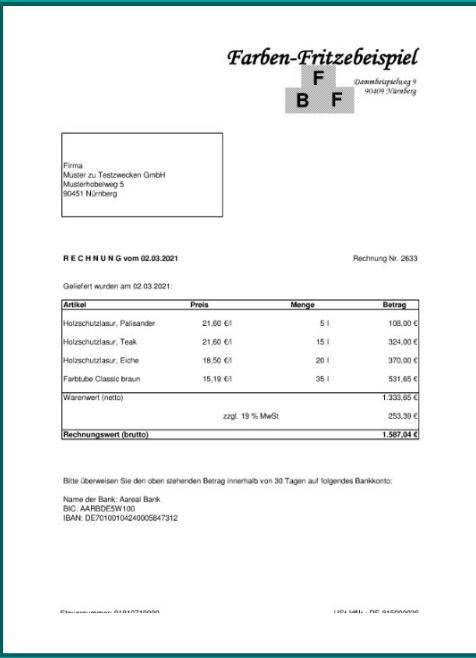


* Source: ICD-Ranking 2020

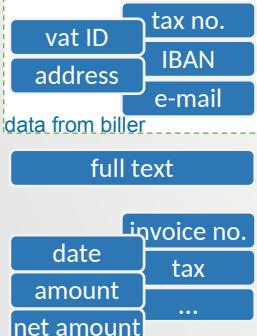
DATEV automation services

(Corporate) Client: uploads document

DATEV Automation Services: provides accounting record proposal



Text
recognition
(OCR) + data
extraction



Accounting record
proposal

- date
- amount
- account
- contra account
- tax
- invoice number
- text
- cost center
- ...

Probabilistic logic modelling in Markov Logic Networks

- ⌚ Expansion of the logical representation (of the 'rules') to incorporate uncertainty by adding probabilities.

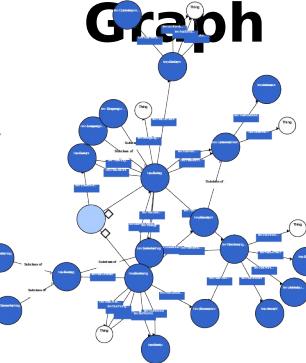
When a carpentry sends an outgoing invoice for furniture, it constitutes revenue. **weight: 5**

When a carpentry sends an outgoing invoice for furniture, it constitutes a fixed asset sale. **weight: 0.4**

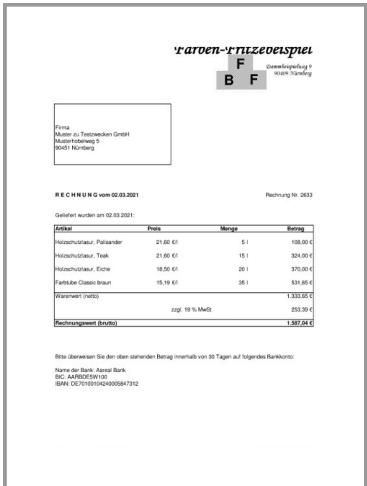


Modelling

Knowledge Graph

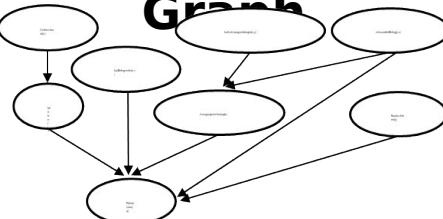


Learning rules



Evidence

Dependency Graph



Inference

Accounting record proposal

- date
- amount
- account
- contra account
- tax
- invoice number
- text
- cost center
- ...



Use Case 2

Geospatial Intelligence

SATCEN



Funded by
the European Union



EUROPEAN UNION SATELLITE CENTRE

SatCen supports the decision making and actions of the European Union in the field of CFSP by providing products and services resulting from the exploitation of relevant space assets and collateral data

EU Agency
located in Madrid
(Spain)

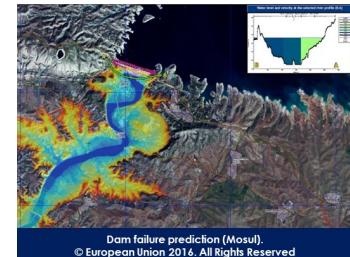
Support to EU Policies and Programmes

Research and Innovation

GEOINT Service
Provider



Support to Evacuation Plan on Nouakchott (Mauritanie)
© European Union 2021. All Rights Reserved



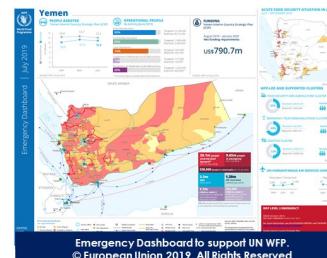
Dam failure prediction (Mosul).
© European Union 2016. All Rights Reserved



Monitoring of blocked maritime traffic during the obstruction of the Suez Canal (2021) / GEO-DAMP Platform
© European Union 2021. All Rights Reserved.



Quick assessment of the damage produced by the explosion in Beirut (August 2020) / GEO-DAMP Platform



Emergency Dashboard to support UN WFP



Distress call detected in Misurata, Libya
© European Union 2011. All Rights Reserved

GEOINT Use Case

“Traditional” GEOINT approaches

- ▽ Change detection
- ▽ Ship/aircrafts/cars detection
- ▽ Identification of roads
- ▽ Surveillance of critical infrastructures
- ▽ Border surveillance
- ▽ Monitoring of illegal activities – deforestation, illegal mining

enexa
technologies

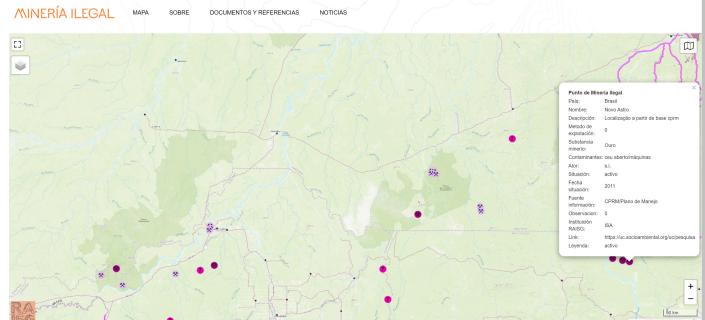
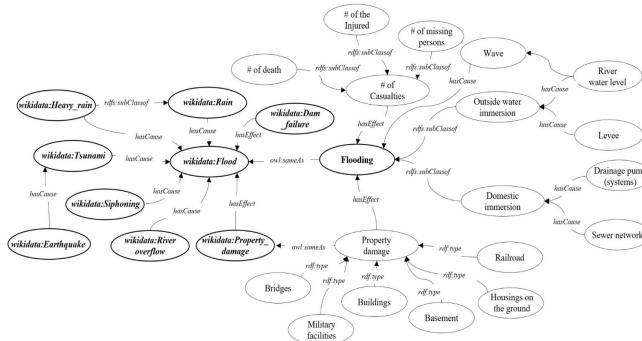


- ▽ Improve detection algorithm
 - ▽ Training with smaller datasets
 - ▽ Better accuracy
 - ▽ Explainability
- ▽ Homogenization and completeness of datasets
- ▽ Distinguish relevant anomalies vs just changes, detection of illegal activities (mining areas), border surveillance, military camps

Exploiting ENEXA technologies

- ▽ **Integration of multiple and heterogeneous sources**
(Homogenization and completeness of datasets)
- ▽ Knowledge Extraction tools
- ▽ Formal Logic for the identification of inconsistencies

- ▽ Application for relevant Change Detection
- ▽ Class expression learning (relevant vs non-relevant)
- ▽ Explainability





Use Case 3

Data-Driven Brand Communication

weblyzard technology



Funded by
the European Union



Publish & Promote

Semantic Search Impact Optimization



Listen & Predict

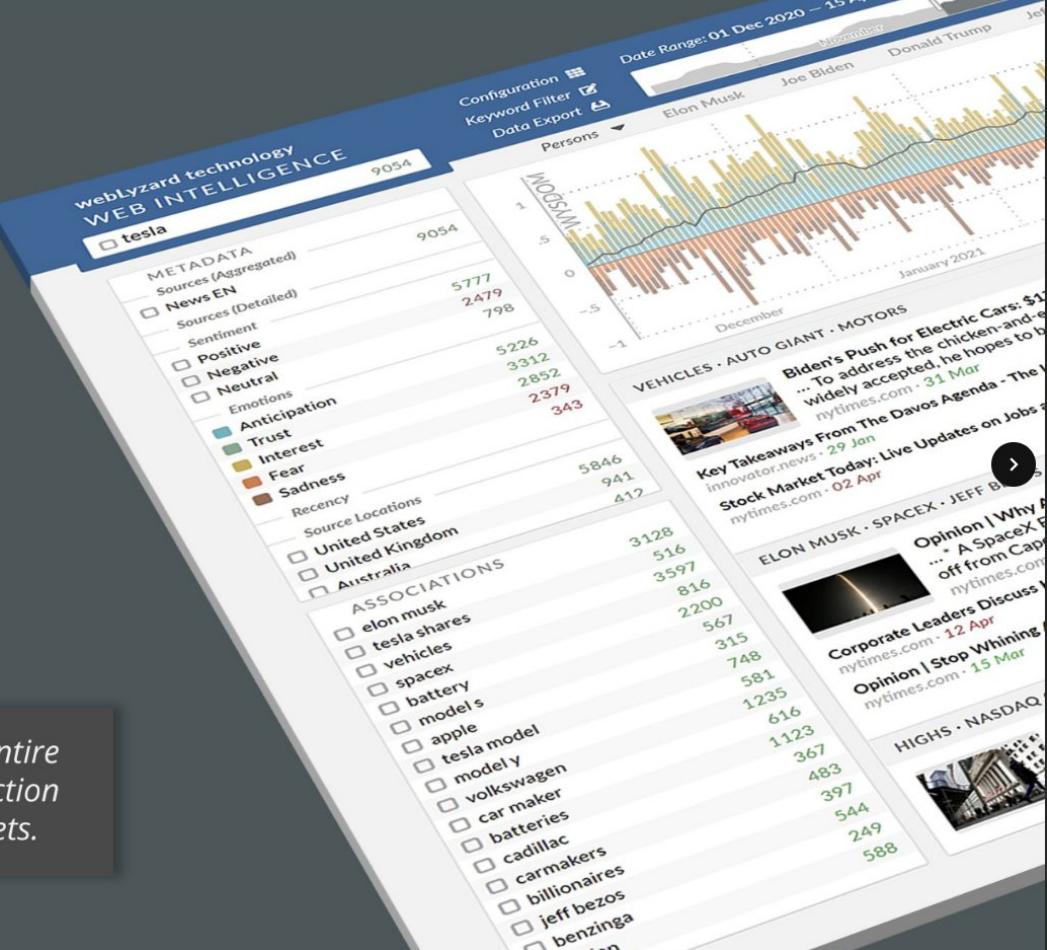
Brand Communication Success Metrics



Integrate & Visualize

Knowledge Hub Visual Analytics

The webLizard web intelligence platform supports the entire **digital content value chain**, from the data-driven production to the distribution, promotion and analysis of content assets.



Funded by
the European Union

Data-Driven Brand Communication

Improve brand communication through data driven techniques

- Gain actionable knowledge from the public debate and stakeholder communication
- Insights for brand communications e.g. marketing campaigns, social media activity, etc..
- See trends, opinion leaders and emerging stories
- Analyse brand perception through sentiment and emotion analysis

ENEXA

- Utilize large webLyzard content collection from News, company, NGO, government websites and social media channels (multiple million documents/month) & webLyzard knowledge graph (KG)
- ENEXA platform brings knowledge extraction & KG embedding components to acquire additional insights on KG entities



Data-Driven Brand Communication

Scenario 1: Suggest Similar Entities

Improve the *Entity Tooltip* of the webLyzard dashboard and suggest similar entities e.g. journalists that report on the same topic, companies that operate in the same industry etc., for collaboration or competition tracking.

- ▷ KG embeddings & vector distance, explainability

Scenario 2: Classification & Information Completion

Provide Metadata for *Source Management* (website and social media sources)

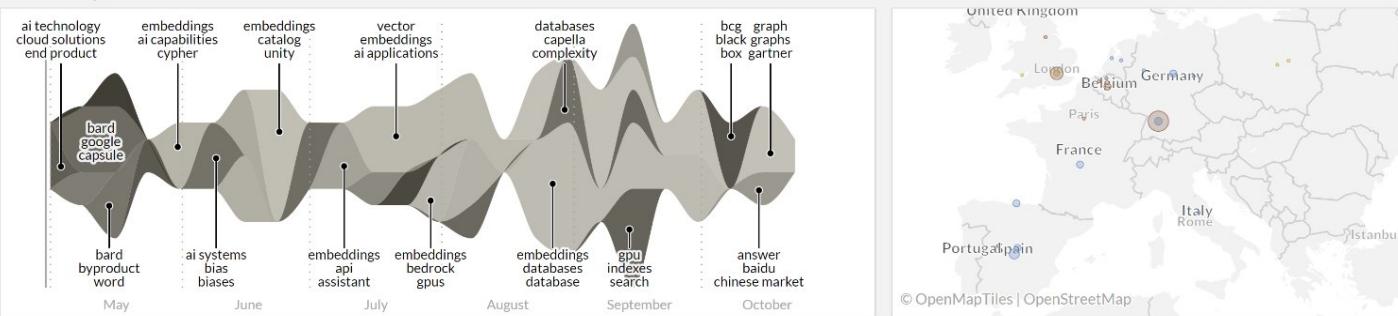
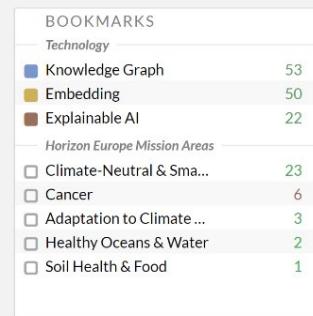
- Deduplicate entries and understand relationships
- Industry classification (NACE codes) for organizations
- Utilize text samples crawled from websites as additional input to identify e.g. a topic/domain

- ▷ Knowledge extraction and KG enrichment, explainability



Large Language Model

7464



Source Classification Filters

METADATA	
Sources - Aggregated	
<input type="checkbox"/> News	4947
<input type="checkbox"/> Companies	2199
<input type="checkbox"/> Non Profit	299
<input type="checkbox"/> Government	19
Sources - Detailed	
Sentiment	
<input type="checkbox"/> Positive	5794
<input type="checkbox"/> Negative	
<input type="checkbox"/> Neutral	
Recency	
NACE Code	

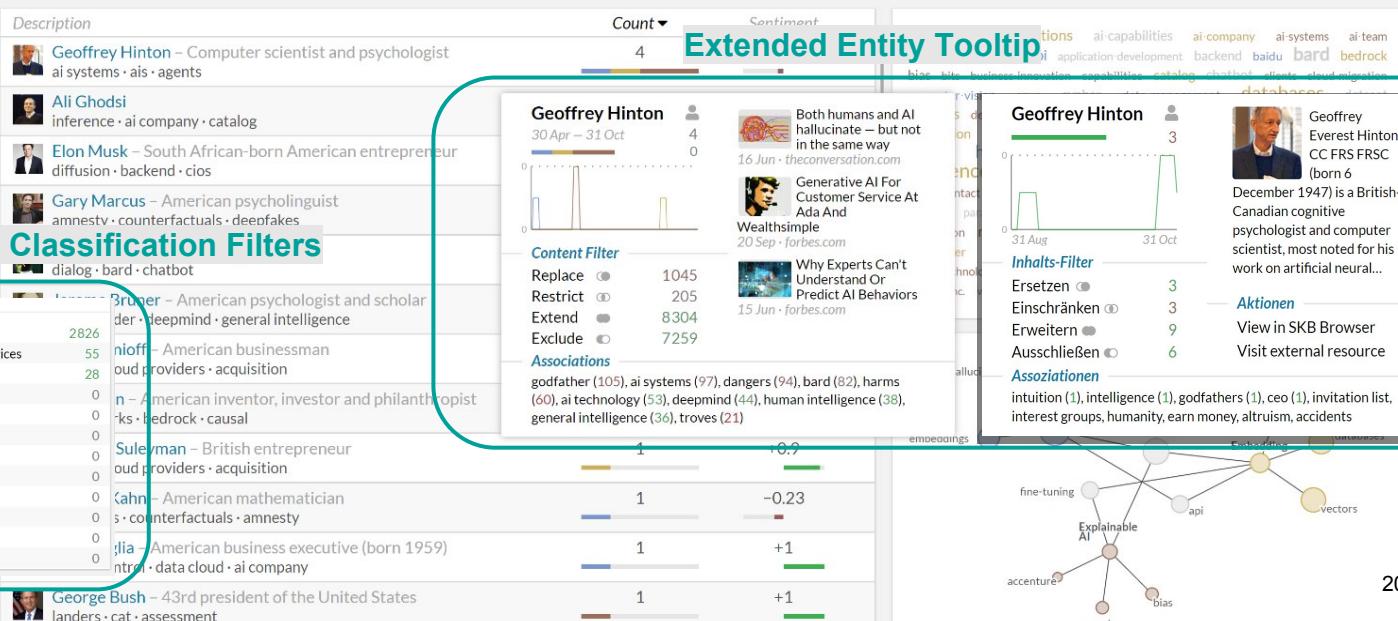
ASSOCIATIONS

- bard
- chatbot
- ai technology
- gpus
- ai systems
- microsoft
- deepmind
- nvidia
- llama
- google
- hallucinations

2222

- J - Information
- M - Professional S
- R - Arts
- L - Real Estate
- S - Other
- P - Education
- Q - Human Health
- B - Mining
- H - Transportation
- E - Water Supply
- G - Wholesale
- A - Agriculture

479



**Thank you
for your attention!**

Contact

DATEV eG

Maria Schnödt-Fuchs

Maria.schnoedt-fuchs@datev.de

Alex Goessmann

alex.goessmann@datev.de

SATCEN

Omar Barrilero – SatCen RTDI Unit

Omar.barrilero@satcen.europa.eu

rtdi@satcen.europa.eu

weblizard technology

Arno Scharl

scharl@weblizard.com

Katinka Böhm

boehm@weblizard.com