

Learning to Predict the Global Risks Interconnections from the Web

[**Minerva:** AI/ML for News]

Dr. Ernesto Diaz-Aviles

Co-Founder, CEO and Chief Scientist at Libre AI
and Adjunct Assistant Professor at UCD

[<ernesto@libreai.com>](mailto:ernesto@libreai.com)

<https://libreai.com>

2018-08-27

Supported by Google through the [Digital News Initiative](#)



Dr. Ernesto Diaz-Aviles
Co-Founder and CEO

Scientist and Engineer with 15 years of experience deploying AI, ML, and Data-driven solutions at scale.

Former Chief Data Scientist and VP at Citi's Innovation Lab. Research Scientist at IBM Research. Research Fellow at the Web Science Lab, L3S Research Center, Germany.



Claudia Orellana-Rodriguez, M.Sc.
Co-Founder and Chief Scientific Officer

Claudia is a scientist and engineer whose work leverages the power of machine learning, natural language processing, social network analysis, and opinion mining to unveil patterns of engagement, attention and influence on the digital era.

Claudia is also a researcher at the Insight Centre for Data Analytics in UCD and collaborator with the MIT Center for Civic Media.



Our mission is to **widely disseminate the benefits** of **Artificial Intelligence** and **Machine Learning** and make them **accessible to the world**

AI and News

We envision a future where **journalists will no longer be limited to report past or current affairs**, but they will be **empowered by Artificial Intelligence to write about future events** with a fair degree of certainty



AI and News

- **Everything is connected** and there are clear historical signs and cycles that produce very similar consequences. The **understanding** of such **interconnections** and causality is fundamental for a comprehensive news coverage
- However, **connecting the dots** and discovering the multiple relationships among events, entities, and global risks are **not trivial tasks for journalists**

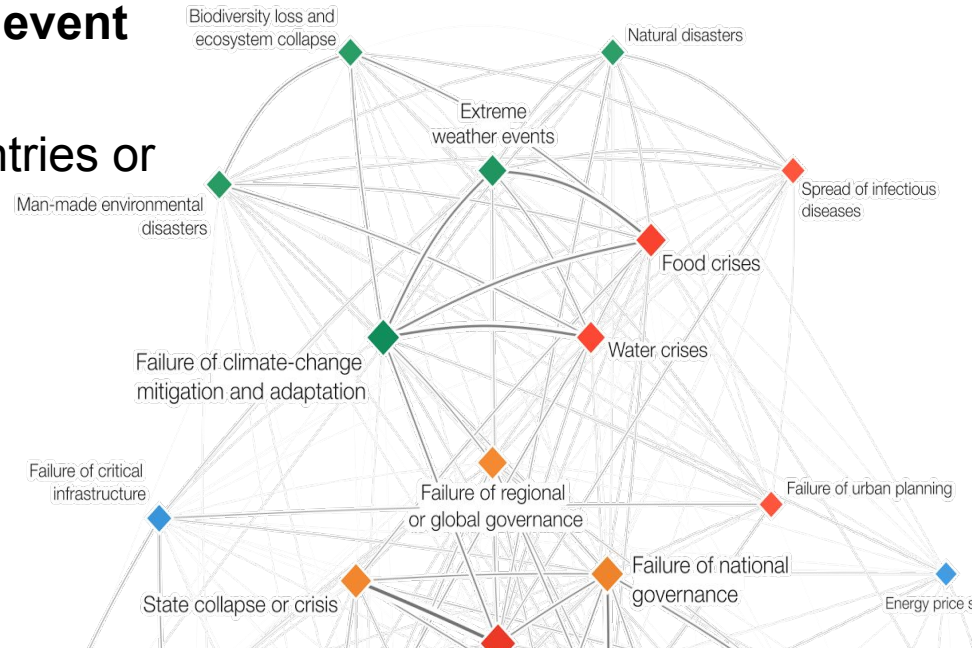


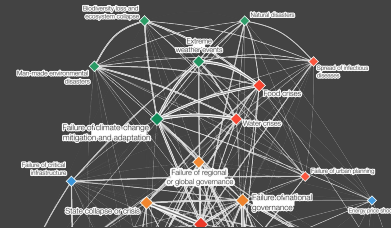
World Economic Forum: The Global Risks Interconnections

A "**global risk**" is defined as an **uncertain event** or **condition** that, if it occurs, can cause significant negative impact for several countries or industries within the next 10 years.

The Global risks 5 broader classes:

- (1) Economic Risks
- (2) Environmental Risks
- (3) Geopolitical Risks
- (4) Societal Risks
- (5) Technological Risks





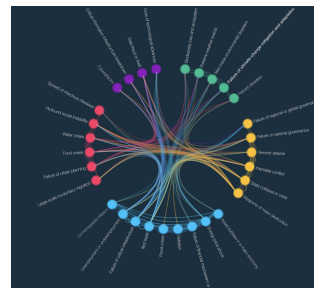
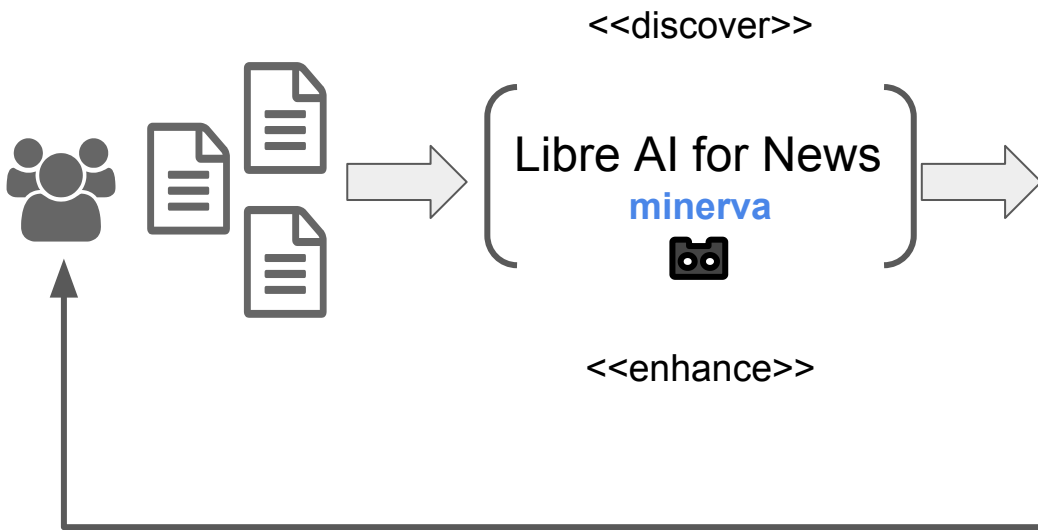
Minerva: automatically generate a
Global Risks Interconnections Map
from **large news datasets** and
web sources

Project: **Minerva**

Learning to Predict the Global Risks Interconnections from the Web

Prototype based on **Artificial Intelligence** and **Machine Learning** that **mines the Web** and **predicts** the (non-obvious) **interconnections** of **global risks** that will be at the core of tomorrow's news

Minerva: Learn to Predict the Global Risks Interconnections from Data



- **Classification** of news articles into Global Risks
- **Detection of key entities:** persons, organizations, locations
- **Unveil existing relationships:** graph of interconnections
- **Predict:** infer future connections



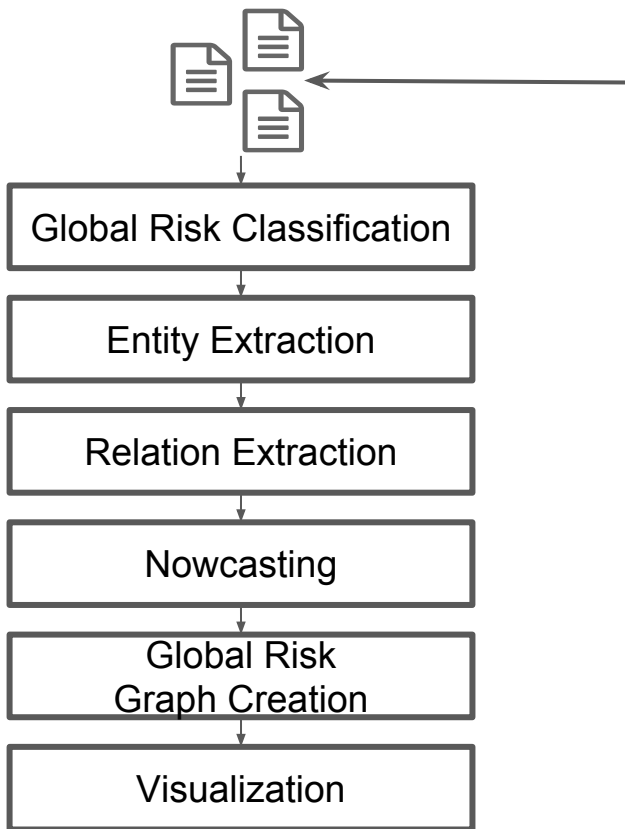
Minerva: AI/ML Pipeline

Common Crawl News

Daily: ~ 4GB - 5GB

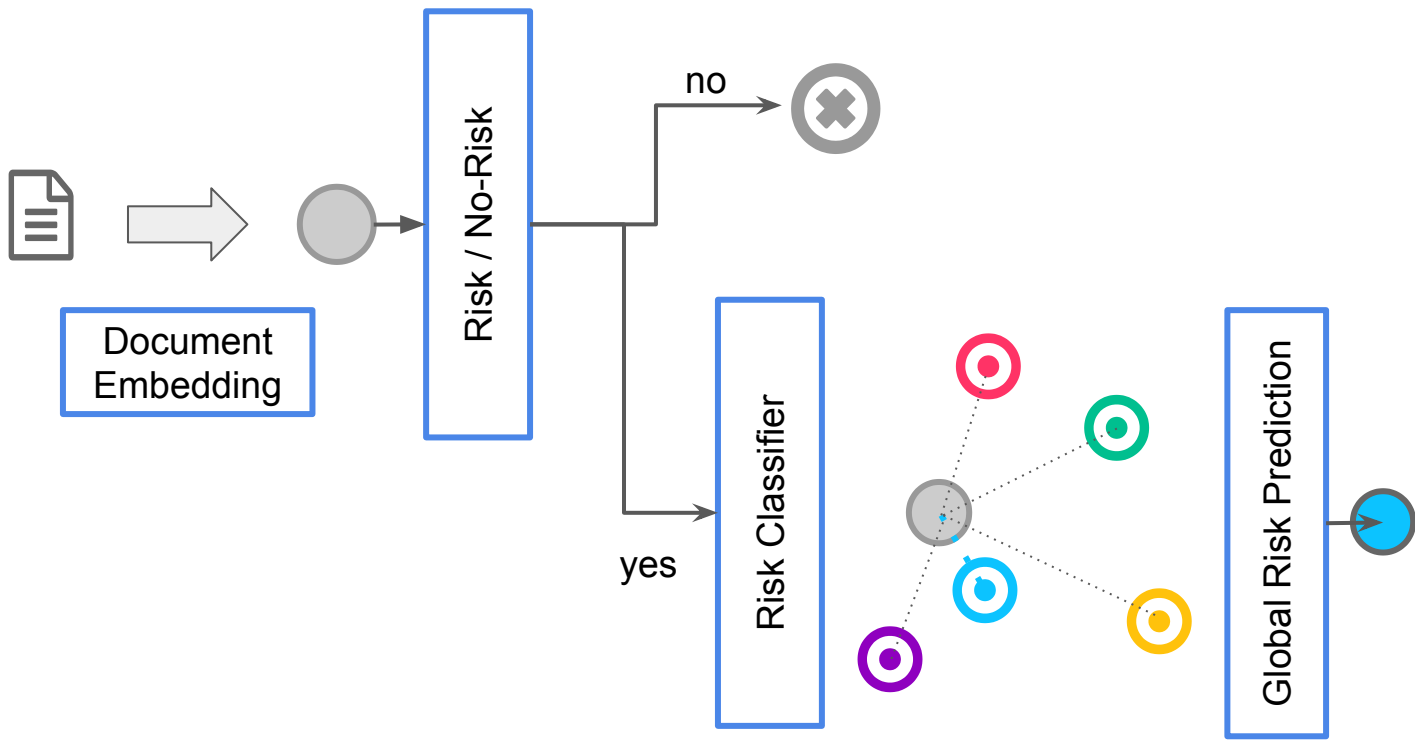
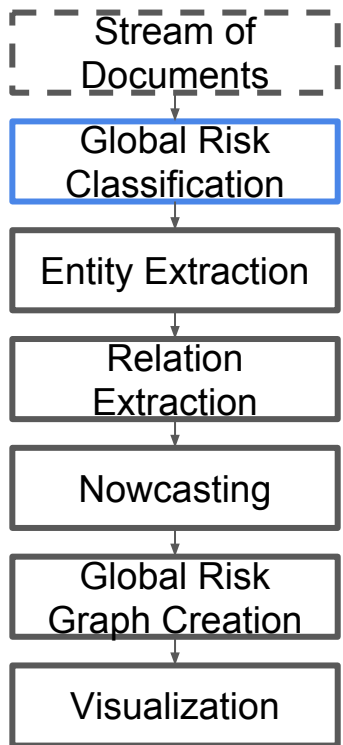
In 2018 ~ 1.5 T

<http://commoncrawl.org/2016/10/news-dataset-available/>



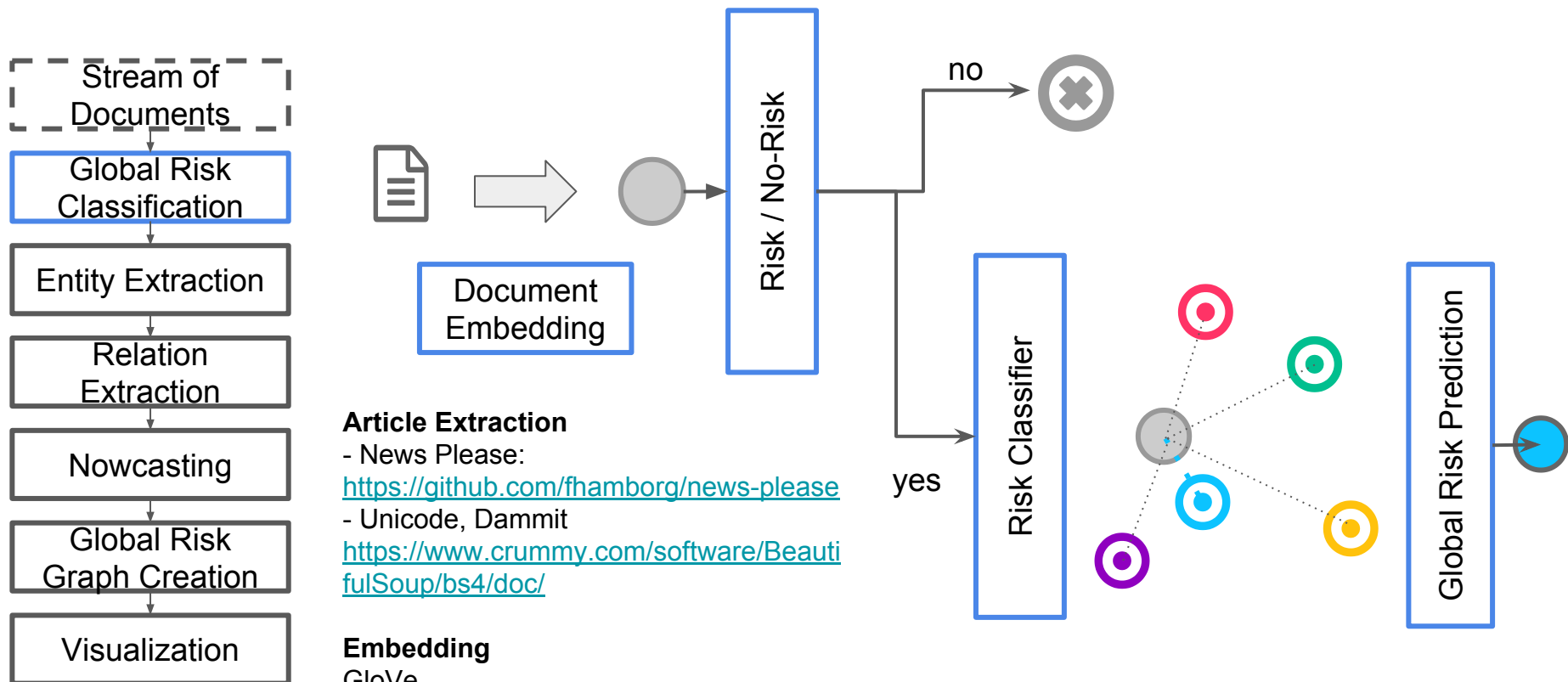


Minerva: AI/ML Pipeline





Minerva: AI/ML Pipeline



Article Extraction

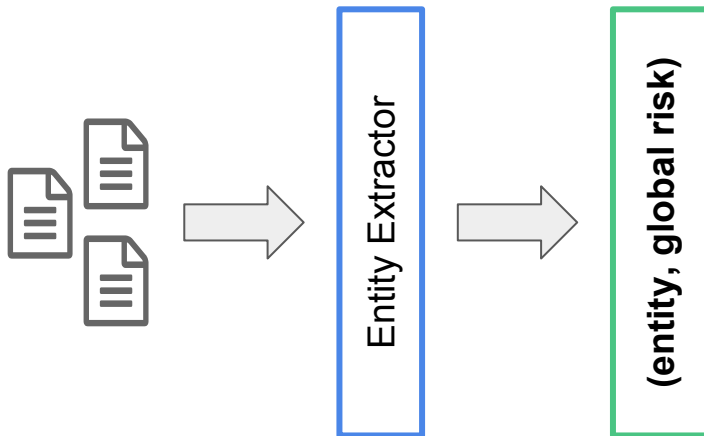
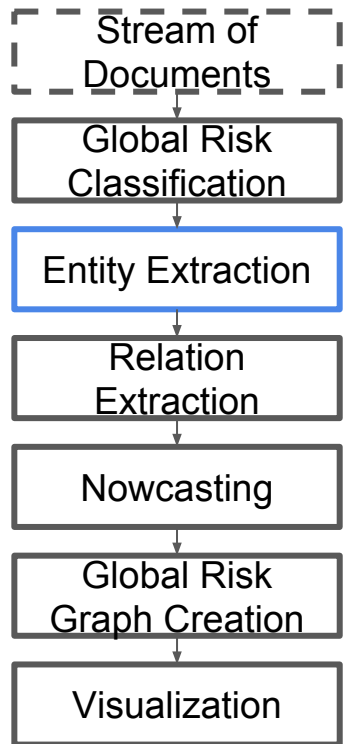
- News Please: <https://github.com/fhamborg/news-please>
- Unicode, Dammit <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

Embedding

- GloVe <https://nlp.stanford.edu/projects/glove/>

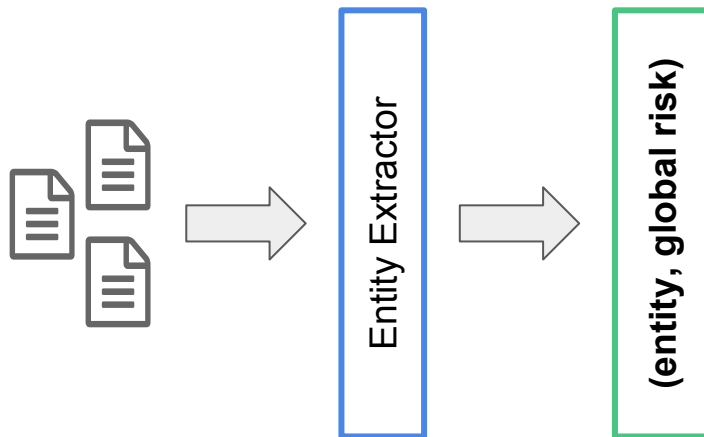
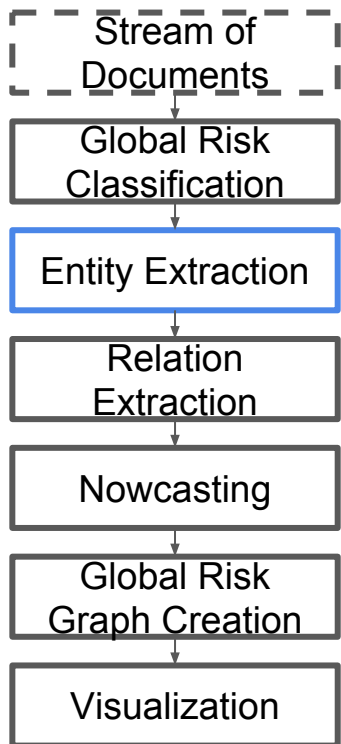


Minerva: AI/ML Pipeline





Minerva: AI/ML Pipeline



NLP - NER

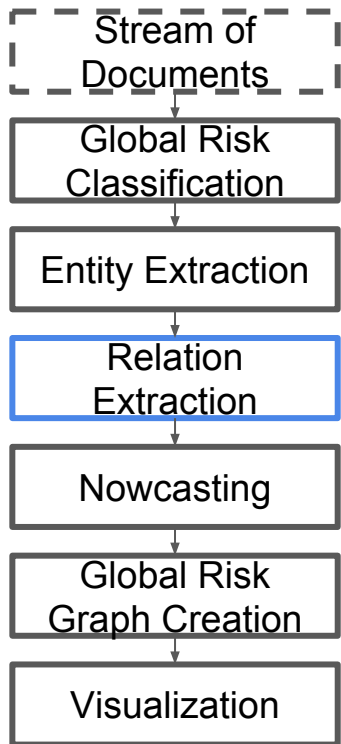
SpaCy: Embed, encode, attend, predict.

CNN + GloVe

<https://spacy.io/>



Minerva: AI/ML Pipeline



(entity, global risk)



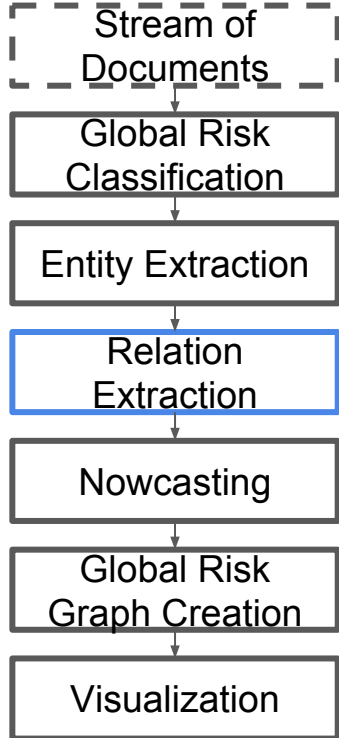
Relation Extraction



Global Risks Links

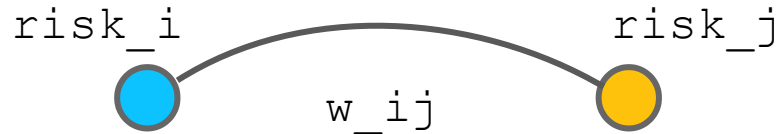


Minerva: AI/ML Pipeline



(entity, global risk)

Relation Extraction



Global Risks Links

Strategy 1: Jaccard Similarity (faster)

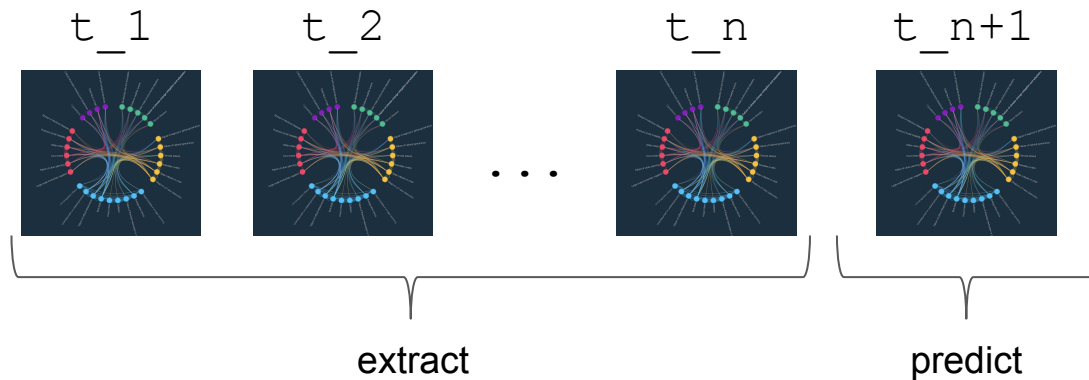
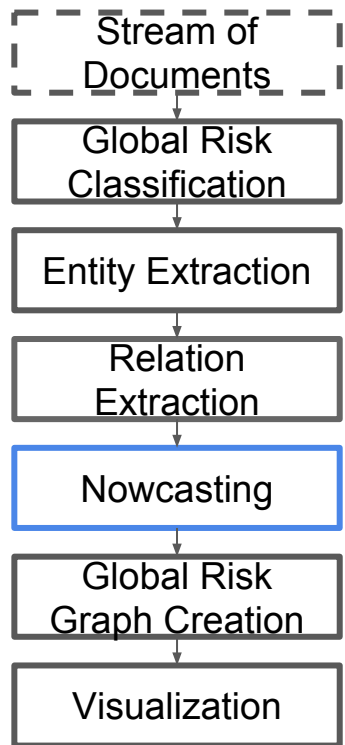
$$w_{ij} = \frac{|entities_i \cap entities_j|}{|entities_i \cup entities_j|}$$

Strategy 2: Semantic Similarity

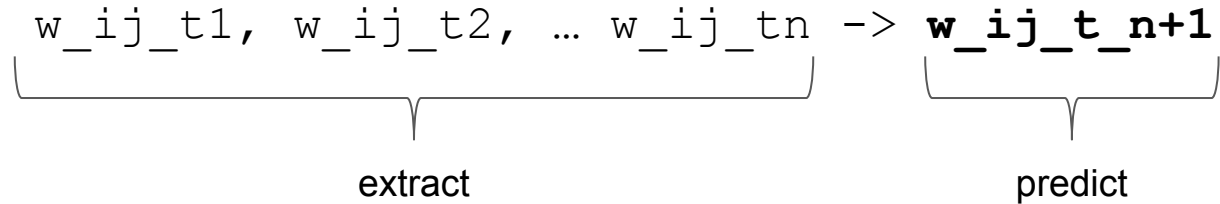
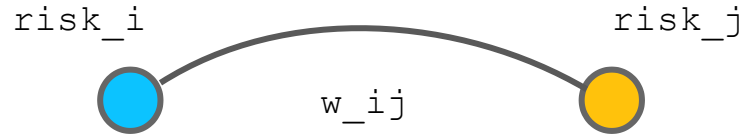
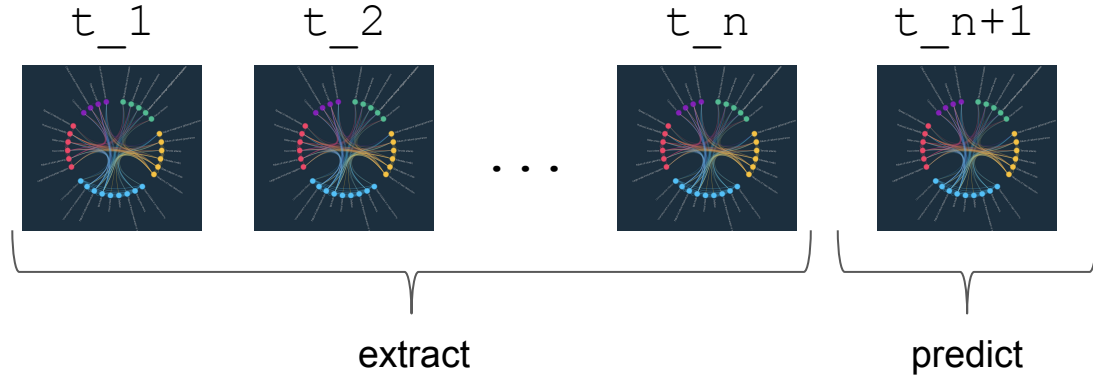
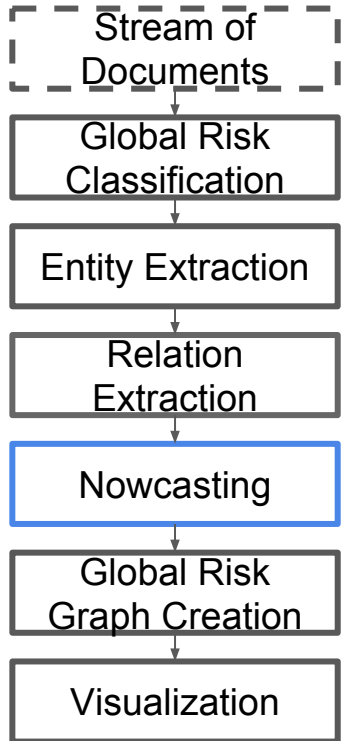
$$w_{ij} = \mathbf{sim}(\mathbf{embedding}(entities_i), \mathbf{embedding}(entities_j))$$



Minerva: AI/ML Pipeline

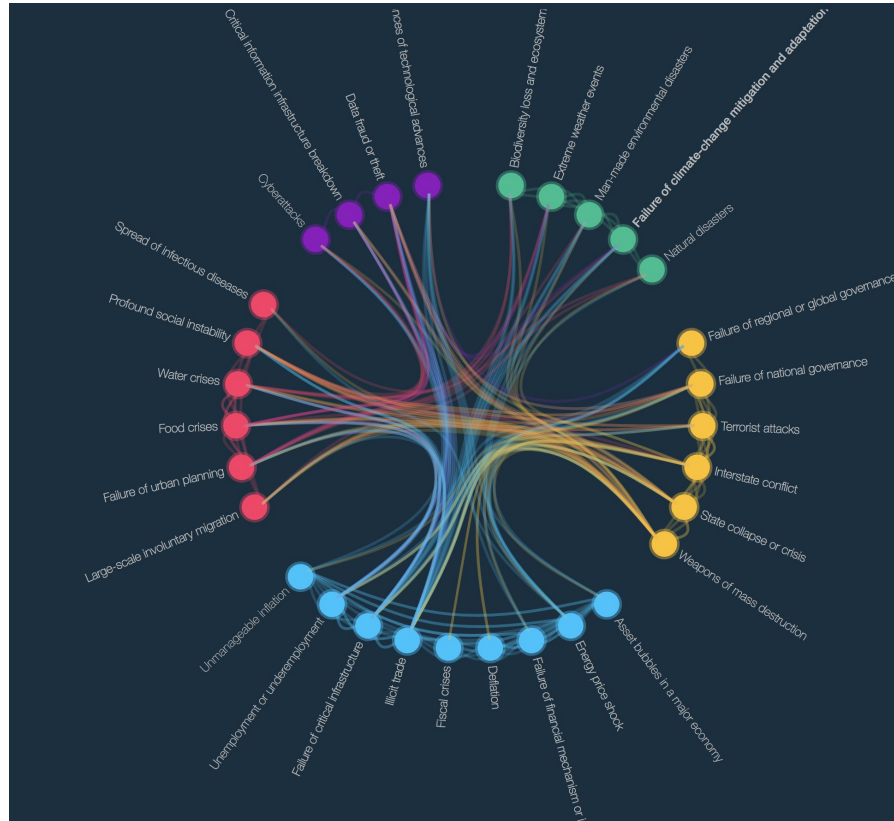
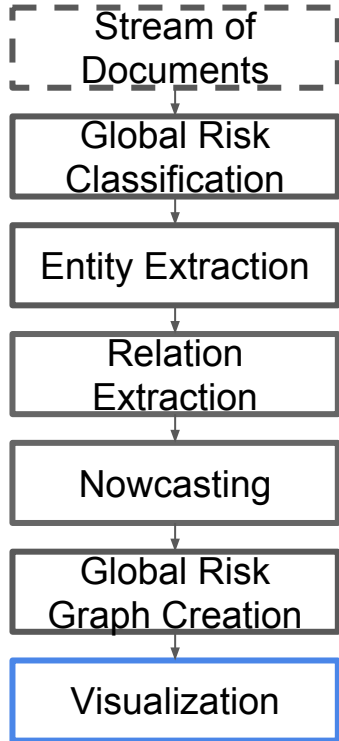


Minerva: AI/ML Pipeline





Minerva: Interactive Graph



Interactive Visualization
D3.js
<https://d3js.org/>



War without Rules

Offensive cyber capabilities are developing more rapidly than our ability to deal with hostile incidents. This creates a fog of **uncertainty in which potential miscalculations could trigger a spiral of retaliatory responses.**

Imagine that a country's **critical infrastructure systems** are compromised by a **cyberattack**, leading to **disruption of essential services** and loss of life—the pressure to retaliate would build rapidly, potentially setting off an escalatory chain reaction. [...]



Experimental Evaluation

- Are the main connections between Global Risks predicted?
- Ground truth: [Web Economic Forum reports](#)
- Metric (averaged over all risks):

$$\text{Precision@n} = \frac{|\{\text{relevant links}\} \cap \{\text{top-n predicted links}\}|}{n}$$

- Dataset **Commons Crawl News 2018** sample (all articles are in English), Articles from Irish News media outlets, Major press agencies, and Major newspapers around the world



Conclusion

- **“It is difficult to make predictions, especially about the future”**
– Danish Proverb
- Predicted Global Risk Interconnection more accurate for short horizons (**Nowcasting**). E.g., months or quarters rather than 10 years as WEF definition
- **Entity based relations** are promising proxy for risk interconnections
- Computationally cheaper **Jaccard similarity leads to better precision** than embedding-based strategy
- Next: continue evaluation and User study
- **Initial rollout with partners. If interested, let me know**



AI and News

We envision a future where **journalists will no longer be limited to report past or current affairs**, but they will be **empowered by Artificial Intelligence to write about future events** with a fair degree of certainty

Thank you!

Dr. Ernesto Diaz-Aviles

Co-Founder, CEO and Chief Scientist at Libre AI
and Adjunct Assistant Professor at UCD

<ernesto@libreai.com>

<https://libreai.com>

LibreAI.com