



X^L Insurance

Tracking Trending Topics in Insurance for Emerging Risk Identification

Amin Karbassi, Ph.D. | Senior Casualty Risk Consultant
Insurance Data Science Conference - London | June 2025

Emerging risks

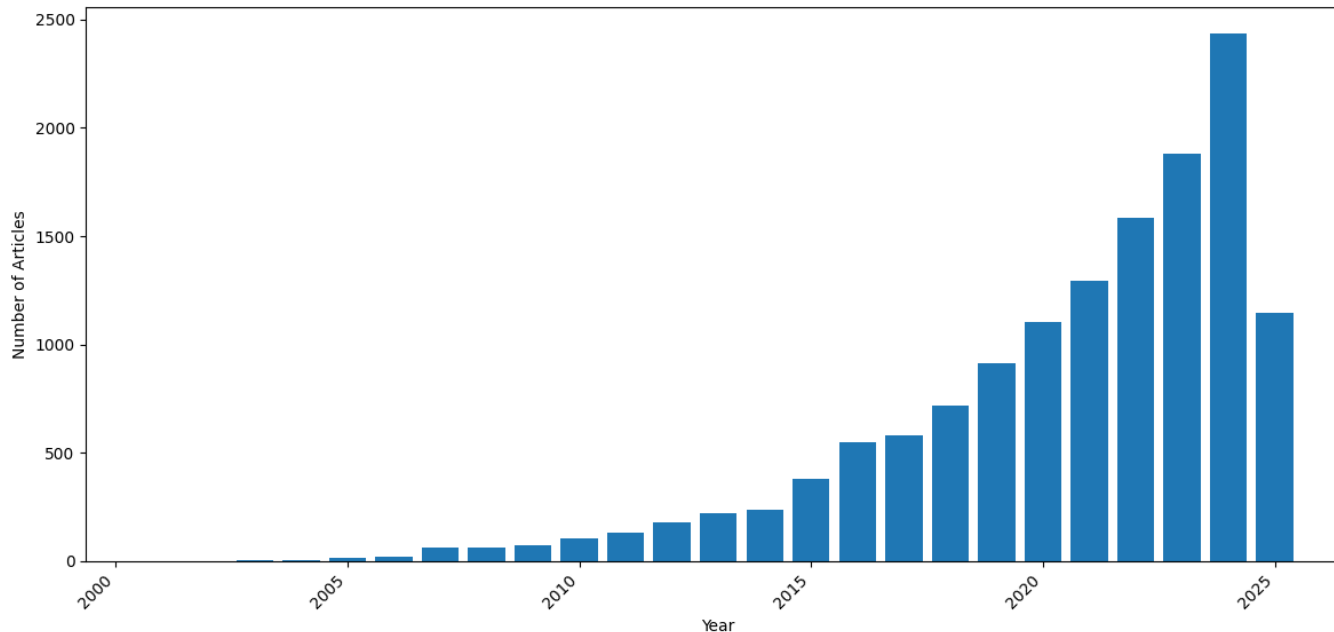
- Newly developing or evolving risks that are not fully understood or quantified
- Often first appear in scientific literature before public recognition
 - PFAS (1990s and early 2000s), Asbestos (1960s), and Opioid Crisis (mid-1990s)
- Traditional clustering methods often ignore time element, limiting trend identification

Goal: Identify **trends** that are not only **thematically coherent** but also **temporally focused**



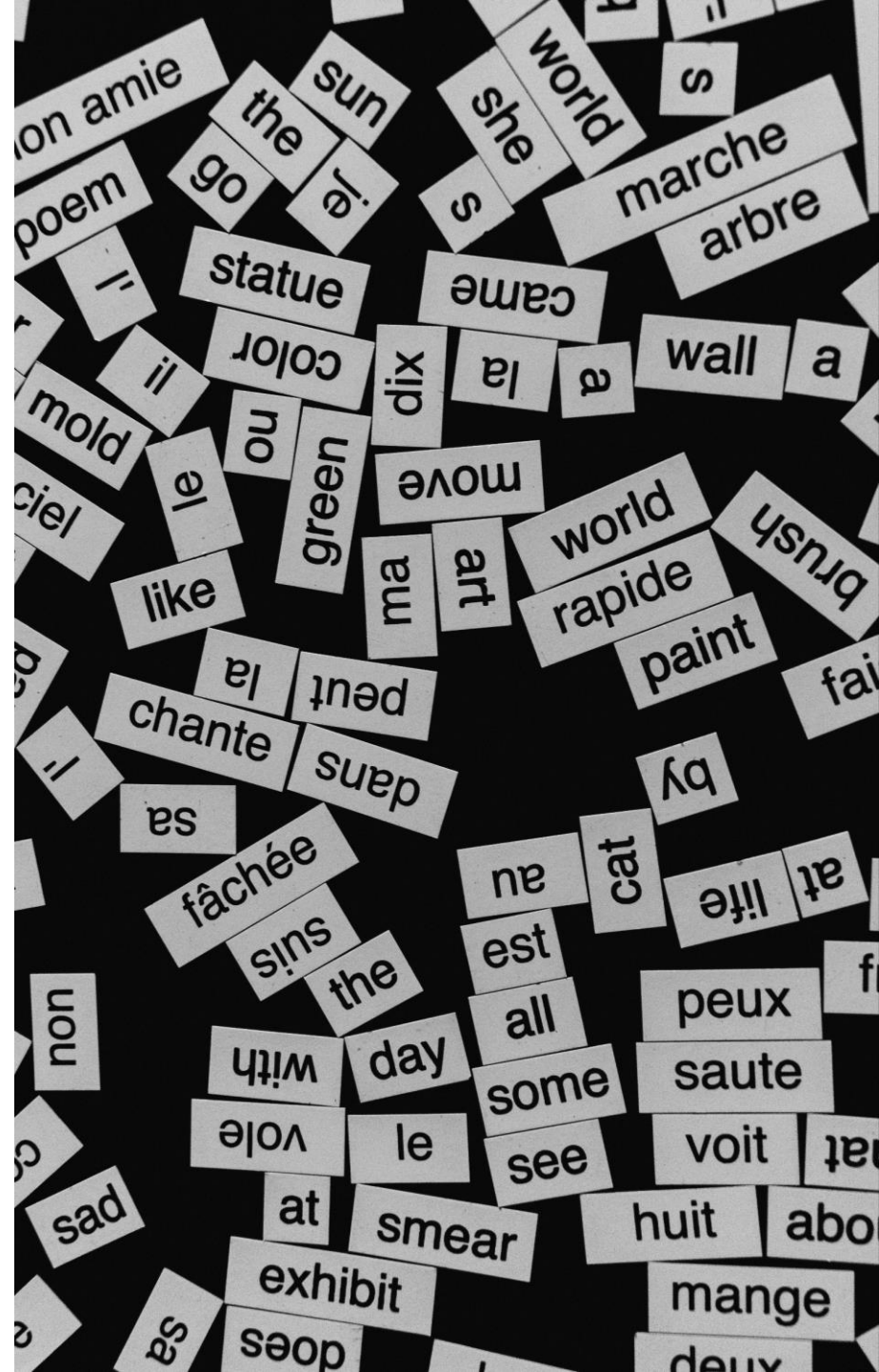
Data preprocessing

- ScienceDirect articles (2000–2025) on **emerging contaminants**
 - 13,705 articles in total
- Tokenization, lowercasing the tokens; removing punctuation, stop words, and numbers; stemming the tokens
 - vocabulary size of over 85,119 distinct words



Text processing

- LDA (Latent Dirichlet Allocation): to derive the terms that carry the main content of the text
 - roughly 40 different features (number of topics) in the entire corpus of all the journal abstracts
- Term Frequency–Inverse Document Frequency (TF–IDF): to compute vector representations for each abstract
 - a matrix of 13,705 rows (one row for each document) and 85,119 columns (one column for each distinct word)
- Singular Value Decomposition (SVD): to reduce dimensionality
 - The vector representations of the abstracts are high dimensional and sparse, thus introduce noise in the clustering
- Temporal (Time-Biased) clustering: to introduce temporal Bias
 - Time (year of publication) is added as a feature prior to clustering



Clustering

- K-means clustering (time-biased)

- **Time Feature Inclusion:** Adds year of publication as a feature in document vectors.
- **Bias Parameter:** Adjusts the weight of the time feature relative to other content features.
- **Weight Adjustment:** Scales the importance of publication year to emphasize temporal grouping.
- **Cluster Localization:** Forms clusters that are both thematically and temporally focused.
- **Trend Detection:** Identifies topic emergence, peaks, and declines within specific periods.



Evaluation

■ Trend Score Metric

- **Silhouette Score:** to measure cluster cohesion; higher is better (range: -1 to 1)
- **Standard Deviation of Years:** to assess temporal spread within clusters - lower SD means documents in a cluster are concentrated around specific time periods

$$\textit{Trend Score} = \frac{\textit{Silhouette score}}{SD_{\textit{years}}}$$

- **Purpose:** Identifies clusters that are both thematically coherent and temporally focused.
- **Outcome:** High Trend Score indicates effective trend detection in specific time frames.



Casualty Risk Consulting

Global Asset Protection Services, LLC, XL Catlin Services SE and their affiliates (“AXA XL Risk Consulting”) provide loss prevention and risk assessment reports and other risk consulting services, as requested. In this respect, our property loss prevention publications, services, and surveys do not address life safety or third party liability issues. This document shall not be construed as indicating the existence or availability under any policy of coverage for any particular type of loss or damage. The provision of any service does not imply that every possible hazard has been identified at a facility or that no other hazards exist. AXA XL Risk Consulting does not assume, and shall have no liability for the control, correction, continuation or modification of any existing conditions or operations. We specifically disclaim any warranty or representation that compliance with any advice or recommendation in any document or other communication will make a facility or operation safe or healthful, or put it in compliance with any standard, code, law, rule or regulation. Save where expressly agreed in writing, AXA XL Risk Consulting and its related and affiliated companies disclaim all liability for loss or damage suffered by any party arising out of or in connection with our services, including indirect or consequential loss or damage, howsoever arising. Any party who chooses to rely in any way on the contents of this document does so at their own risk.

AXA, the AXA and XL logos are trademarks of AXA SA or its affiliates. © 2025.



Know You Can