Risk Management (2026) 28:28
https://doi.org/10.1057/s41283-026-00214-x

RESEARCH

A model for high-frequency detection of current risks based on news analysis and decentralized social networks

Timotej Jagrič[1] · Maša Galun[1]

Received: 12 February 2026 / Accepted: 12 April 2026

Abstract

This article presents a data-driven model for identifying current risks in non-financial companies, supporting managerial and supervisory decision-making. The model integrates signals from the global news database GDELT and the decentralized social media platform Nostr into a structured risk register aligned with COSO ERM and ISO 31000 frameworks. Using standardized indicators of relevance and dynamics, it enables comparable monitoring of changes in the risk environment over time and across different sources. Empirical results from both sources indicate remarkable differences while complementing each other: news data provides a more stable, event-driven signal, while decentralized signals enable earlier risk detection. Forward-looking validation demonstrates that these indicators contain economically meaningful information, improving the prediction of future systematic risk beyond traditional approaches. The model does not replace existing methods of risk management, but supplements them with continuous, data-driven input into the decision-making processes of management and supervisory boards.

Keywords News analysis · Social media · Risk register · Risk management

Mathematics Subject Classification D81 · G32 · C55

✉ Maša Galun
masa.galun@um.si

Timotej Jagrič
timotej.jagric@um.si

1 Faculty of Economics and Business, Institute of Finance and Artificial Intelligence, University of Maribor, Maribor, Slovenia

Published online: 20 April 2026

Introduction

Nonfinancial companies operate in an environment of increasing complexity, rapid propagation of shocks, and high interdependence among economic, technological, regulatory, and geopolitical factors. Classic risk management approaches in companies (EMR), which are mostly based on periodic quantitative evaluations and internal reports, are increasingly struggling to keep up with the dynamic nature of the external environment. This is especially evident in risks such as production chain disruption, cybersecurity, regulatory risks, geopolitical stability, and technological disruption, which often manifest first in the information space and only later in real business indicators.

In the last two decades, the use of large text corpora as a source for quantifying uncertainty, risk, and expectations has become established in economic and financial literature. Baker et al. (2016) showed that it is possible to create an index of economic-political uncertainty that shows a statistically significant link with investment activity, employment, and financial volatility. Later research extended this approach to geopolitical risks (Caldara and Iacoviello 2022) and to individual-company risks, using transcripts of conference calls (Hassan et al. 2019). Other studies further demonstrate that textual information from news and financial disclosures can provide early signals of financial and operational risks and improve the predictive performance of risk assessment models (Groth and Muntermann 2011; Barakat et al. 2018; Huang and Lim 2024). More recent research also applies large-scale text mining and machine learning techniques to identify risk signals in financial news and other digital sources, enabling systematic monitoring of corporate and macroeconomic risks (Roeder et al. 2022; Chen and Ji 2025; Pei et al. 2025). The common denominator of these studies is the assumption that risks are first reflected in discourse (in what is being discussed, how intensely, and what dynamics) and only then in real economic outcomes.

The contribution of this article is the development and analysis of a model for detecting emerging risks in the non-financial sector, which extends existing one-dimensional uncertainty indices and introduces a multidimensional risk-detection model. The model is based on the conceptual design presented in Jagrič et al. (2025), using the GDELT global news database and an innovative AI model for risk register design. In this article, we upgrade this model by using decentralized social media Nostr. Both sources are analyzed with the same risk register, which is structured in accordance with established risk management frameworks (COSO ERM and ISO 31000), enabling direct integration in the risk management process.

The main thesis of this article is that combining news and decentralized social media signals enables earlier and richer identification of emerging risks than either source could provide on its own. In addition, the GDELT database provides a relatively stable and widely representative signal, whereas Nostr is highly volatile, sensitive to early changes in discourse and niche communities. The comparison analysis of both sources reveals areas of consensus and systematic discrepancies with important implications in risk interpretation.

The model presented in this article not only detects emerging risks from heterogeneous textual sources, such as GDELT and Nostr, but also provides forward-looking validation by linking these signals to future changes in systematic risk (Jagrič et al. 2026). This demonstrates that structured, high-frequency textual indicators carry predictive information that can enhance managerial and supervisory decision-making, complementing traditional risk assessment processes. Designed as a decision-making tool, the model enables managers and supervisory boards to monitor changes in the external risk environment and integrate heterogeneous information signals within structured risk management processes. The article contributes on multiple levels: it develops a multidimensional risk detection model that exceeds aggregate uncertainty indices and enables analysis at the level of individual risks; it combines news and decentralized social media sources within a single methodological framework for the first time; and it explicitly connects empirical risk monitoring with normative corporate risk management frameworks, allowing direct practical application.

The second chapter provides an overview of relevant scientific literature on textual risk measurement, the use of news databases, and the role of decentralized social media. The third chapter briefly presents the databases and their statistical properties. The fourth chapter is devoted to the risk register, its structure, and its connection with COSO and ISO standards. The fifth chapter develops the methodology and formally defines all the indices used. In the sixth chapter, the empirical results and comparative analysis are presented. Chapter seven concludes the article with implications for practice and economic policy.

Literature review

The use of text sources for quantifying uncertainty and risk has a long tradition in economics, but only with the development of computing capabilities and large digital corpora was the systematic, repeatable empirical application of these approaches possible. Groth and Muntermann (2011) showed that information from news sources can be used to detect sudden changes that are often overlooked by classic approaches. Text mining and machine learning techniques enable the identification of patterns that provide early signals of abnormal market events. Barakat et al. (2018) also emphasize that textual sources provide important additional information regarding financial, strategic, and operational risks. This theory is further supported by the results of Huang and Lim (2024), who analyzed the integration of financial indicators with textual data for financial risk management. Their study shows that incorporating textual information can significantly improve risk management models' ability to detect potential financial risks.

The literature also presents many multidimensional models for early detection of financial risks that combine financial statements, market data, and textual information. The proposed models can detect risk several months before it occurs. Studies also emphasize that the best predictive performance is achieved by combining multiple indicators, and that this approach outperforms classical statistical models and standard machine learning models (Chen and Ji 2025; Roeder et al. 2022).

The use of textual data for financial risk analysis is further explored by Pei et al. (2025), who develop a framework to identify company-related risks from financial news. Their approach classifies risk factors into multiple categories and demonstrates that transformer models can effectively detect risk signals in large collections of news articles, enabling large-scale monitoring of corporate and macroeconomic risks. The role of textual information in financial risk assessment is further highlighted by Stander (2024), who develops a news sentiment index that can serve as an early indicator of systemic and credit risk under IFRS 9 standard. A systemic index, based on news analysis, was also developed by Ma et al. (2025), who showed that their index successfully predicts macroeconomic changes and serves as an early warning tool for systemic risk monitoring.

Baker et al. (2016) demonstrated through an analysis of the frequency of specific terms in newspapers that it's possible to construct a time series of economic-political uncertainties that statistically significantly correlates with macroeconomic variables. Their methodology has set a standard for later research, as it's based on a clear definition of key words, transparent index construction, and empirical validation. Caldara and Iacoviello (2022) have enhanced this approach, focusing on geopolitical risks. Their analysis shows that geopolitical shocks, measured with text indicators, have a significant impact on investments, trade, and financial markets. The key message of this research is that text indicators don't measure objective risks, but rather perceived and communicated risk, which is precisely the channel through which expectations and decisions are transmitted to the economy.

A major movement in the literature is the work of Hassan et al. (2019), which shifts risk from the macro level to the level of individual companies. The authors show that political risk is highly heterogeneous among companies and that aggregate macro indicators mask significant differences in exposure. This directly supports the approach of this article, which does not treat risk as a single index, but rather as a multidimensional vector, structured in a register.

Parallelly, a large news database was developed, with GDELT occupying a central place. Hopp et al. (2019) emphasized that the main advantage of GDELT is its openness and scope, but they also point out methodological challenges related to deduplication, heterogeneity of sources, and the interpretation of metadata. A recent study by Hong et al. (2025) empirically showed that some systematic errors in GDELT can affect results if not properly addressed.

Relatively new but rapidly growing, a branch of literature deals with decentralized social media. Raman et al. (2019) analyzed federated networks and showed that decentralization brings specific structural properties that influence the dissemination of information. Wei and Tyson (2025) are the first to systematically analyze Nostr and show that its decentralized architecture leads to distinct dynamics in the availability and replication of content, with direct consequences for signal measurement.

The model presented in this article conceptually enhances the approach developed by Jagrič et al. (2025), which operationalized risk detection for the first time using a global news database. While the previous analysis was focused on a single news source and aggregated topicality indicators, this article extends the methodological framework by including a decentralized social media platform and enables a systematic comparison of heterogeneous information signals within a single register of risks.

The existing literature offers either aggregate risk indicators or analyses of individual sources, either aggregate risk indicators or analyses of individual media sources, rather than an integrated model that would enable the simultaneous analysis of multiple heterogeneous information flows and their direct mapping to the risk register, consistent with corporate risk management practices. In comparison to existing literature, this article combines three research flows:

- Textual risk measurement
- The use of large databases
- Analysis of a decentralized social media platform

The key innovation is the integration of flows in an operative model, which is directly related to risk management in companies.

Database

The empirical analysis in this article is based on two data sources that are remarkably different in their nature, structure, and information dynamics, but offer complementary risk detection. The first source is the Global Database of Events, Language, and Tone (GDELT), which aggregates global news and structures it as time- and theme-coded events (The GDELT Project, 2022). The second source presents the decentralized social media platform Nostr, in which content is distributed via a network of relay servers without a central administrator (Nostr Protocol Developers, 2023). Both sources are analyzed through a single risk register, enabling direct comparison of signals within the same conceptual space.

In the empirical part, the results are calculated from data up to December 2025, generated by automated capture and processing scripts. Data are organized in a panel, where each observation is a pair (risk, month). This structure enables the use of standard statistical approaches for analyzing time series and cross-sections, as well as the direct comparison of the dynamics of individual risks over time.

The time coverage of both sources differs: GDELT covers a longer historical period (from January 2017 to December 2025), spanning 108 consecutive months. Nostr, on the other hand, covers the period from January 2024 to December 2025, a total of 24 months (due to technical limitations). The difference in time-series horizons has direct consequences for the stability of statistical estimates, especially for indicators based on historical distributions. Nevertheless, all indicators are defined using sliding windows, which enables methodological consistency across all time series.

Table 1 presents the basic properties of both databases, including time coverage, the number of risks, and the number of observations. Nostr panel contains 6336 observations, which over 24 months implies a total of 264 risks, present in all time intervals. The GDELT panel contains 19,548 observations, which over 108 months amount to 181 risks. In both cases, the panels lack no observations, which makes statistical analysis and interpretation of results much easier.

The basic variable *value* presents the intensity of individual risk detection in a given month. At Nostr, this variable is expressed as the absolute number of detected posts, whereas at GDELT it is already normalized and expressed as a relative intensity of news reporting. Because of this difference, the values of both sources are

Table 1 Basic characteristics of databases

Source	Nostr	GDELT
n_obs	6336	19,548
n_months	24	108
n_risks	264	181
Start	1.01.2024	1.01.2017
End	1.12.2025	1.12.2025
panel_density	1	1
monthly_total_mean	26,852.75	0.51144
monthly_total_sd	12,909.34	0.109384
monthly_total_cv	0.480746	0.213874
avg_top20_share	0.728208	0.894555
avg_top50_share	0.917922	0.98747
gini_total_by_risk	0.849504	0.904467

Source: Global Database of Events, Language, and Tone (GDELT), Notes and Other Stuff Transmitted by Relays (Nostr)

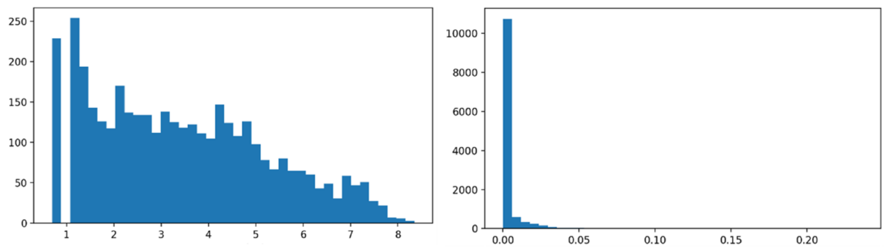

Fig. 1 Distribution of the Variable *value*. *Source*: Author's own calculations. *Note*: The x-axis represents , and the y-axis shows frequency. The histogram on the left displays the results for Nostr, while the histogram on the right shows the results for GDELT

not directly comparable, which further justifies the use of relative and standardized indicators in further analysis.

The distribution of the variable *value* shows a distinctly asymmetrical structure with a long tail in both sources. A proportion of zero observations at Nostr is around 40%, indicating that a large amount of risks in a given month is not detected in network discourse. A similar, but even more intense concentration is present in GDELT. On the normalized scale, the absolute values are significantly smaller, but the distribution's structure is comparable. A proportion of zero observations is also high in GDELT, which confirms that, in each month, most risks don't reach a noticeable threshold in the news flow (Fig. 1).

The analysis of aggregated activity by months further highlights the differences between the sources. With Nostr, the average monthly news volume is approximately 26,853, and the standard deviation of monthly aggregates does not exceed 12,900, which makes the covariance coefficient close to 0.5. This volatility reflects the impulsive nature of decentralized social media, where the discussions about risks often occur in short, intense waves. With GDELT, the average monthly aggregated value is around 0.51, and the standard deviation is around 0.11, giving us a coefficient of variation of around 0.2. That makes the news flow significantly more stable and inertial (Fig. 2).

The concentration of risk activity is extremely high in both sources. If we classify risks by total volume over the entire period, we see that the first ten risks in the Nostr account for about half of all activity. This amount is even higher in GDELT's activity (about 75% of all activity). Formally, this concentration is captured by the Lorenz curve and the Gini coefficient, which is approximately 0.85 for Nostr and exceeds 0.90 for GDELT. This means that a small number of risks systematically dominate the information space, while most remain in the background.

These results have some important methodological implications. Firstly, the absolute volume of appearances is not an appropriate indicator of topicality, as it is heavily biased in favor of structurally dominant risks. Secondly, standardization based on each individual's risk history is essential to detect relevant changes for risk management. Thirdly, differences in volatility between sources confirm that Nostr and GDELT do not measure the same phenomenon but rather two distinct dimensions of risk perception, further justifying their simultaneous use within a multidimensional model of risk detection.

The risk register as a central structural element of the model

The risk register represents the conceptual and operational core of the model for the detection of current risks, presented in this article. Its role is to establish a systematic and repeatable mapping between raw text data from the external environment and a structured risk management framework. Unlike traditional risk registers, which are often a result of internal workshops or expert assessments and are updated periodi-

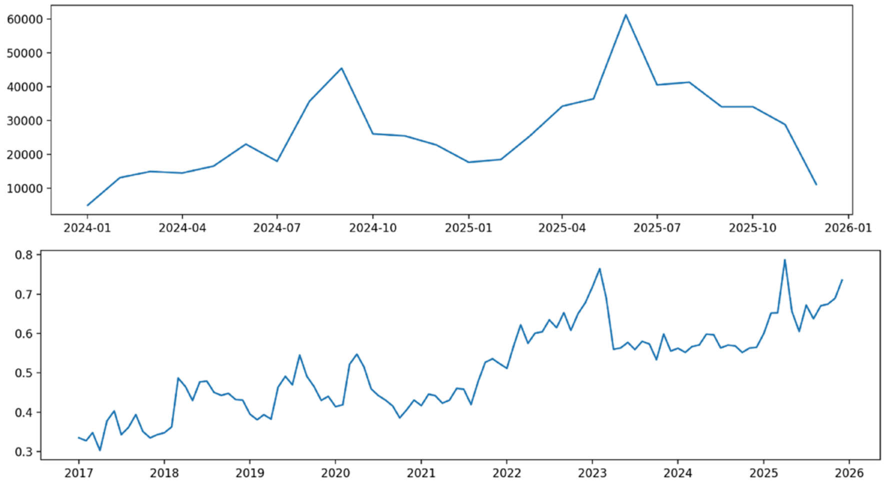

Fig. 2 Monthly Total Volume (Sum Across all Risks). *Source*: Author's own calculations. *Note*: The x-axis shows time (monthly intervals), and the y-axis shows the number of detected posts. Results for Nostr are presented above, and results for GDELT are shown below

cally, the register in this model is designed as a dynamic taxonomy that enables continuous detection and quantification of external risk signals.

The risk register mentioned is structured in a two-tier hierarchy, consisting of risk groups and individual risks. This design is not random; it reflects a conscious methodological decision to ensure comparability across content domains and to limit structural bias that could arise from excessive fragmentation of some areas at the expense of others.

The register consists of 22 content groups (*risk_group*) that represent fundamental risk domains, relevant for non-financial companies. Each group consists of 12 individual risks, for a total of 264. This balanced structure ensures that no risk group is not a priori more represented in the analysis solely because of the larger number of defined risks, which is important for later aggregations and comparisons between groups.

Each individual risk is identified by a unique risk code (*risk_code*), described by an English name (*risk_name_en*) and a short description (*description*). In addition to that, each risk is associated with a set of eight semantic keywords (*keyword*) that operationalize that risk in the text space. Altogether, the register contains 2112 lines (a product of 264 risks and 8 keywords per risk). The current implementation of the registry focuses on the English-speaking information space.

An important feature of the register is that it does not contain explicit subgroups of risks. That means that the mentioned hierarchy is limited to two levels: a group and an individual risk. This design is consistent with the model's purpose, as it enables sufficiently granular analysis to identify specific risks while maintaining transparency and the integrability of results at the group level.

The mapping of external text sources to the risk register is based on the semantic operationalization of individual risks through predefined key phrases. Each risk is therefore defined not only conceptually, but also operationally, as a set of terms whose appearance in the text signals the potential presence or actuality of that risk. Formally, we can describe the risk register as a set of triples:

$$\mathcal{R} = \{(g_k, r_k, \mathcal{W}_k) \,|k = 1, \ldots, 264\},$$

where g_k denotes the risk group, r_k denotes an individual risk and $\mathcal{W}_k = \{w_{k_1}, \ldots, w_{k_8}\}$ is a set of all keywords, associated with the risk r_k. Let $D_t = \left\{d_{t_1}, \ldots, d_{t_{N_t}}\right\}$ be a set of all documents (news or posts) in a month t. The document d_{t_i} is assigned to risk r_k, if and only if

$$\exists w \in \mathcal{W}_k : w \in d_{t_i}.$$

Based on this, the monthly volume of detections of risk r_k in a month t is defined as

$$V_{k,t} = \sum_{i=1}^{N_t} \mathbf{1}(d_{t_i} \to r_k),$$

where $\mathbf{1}(\cdot)$ is the indicator function. This formalization ensures complete transparency of the mapping and ensures full reproducibility of the analysis, which is crucial for scientific use and practical implementation in businesses.

The COSO Enterprise Risk Management (2017) framework emphasizes that risk identification must be systematic, structured, and integrated in strategic decision-making. The risk register in this model directly supports the risk identification component, enabling continuous detection of external factors that can influence the achievement of the company's strategic goals. The individual risks in the risk register can be understood as standardized risk statements that are comparable over time and between companies.

The balanced structure of the register further supports COSO's requirement for a comprehensive view of risks. Since each group has the same number of risks, the aggregated indicators at the group level reflect not only differences in granularity but also actual differences in the perceived relevance of risks in the environment. The developed indicators can thus be directly used as inputs to probability assessment processes and risk prioritization, without replacing the professional assessment of risk bearers.

The ISO 31000:2018 standard requires that risk management is based on the best available information, that it is structured, and that it is implemented on a regular basis (International Organization for Standardization, 2018) The risk register, used in this model, meets these requirements on several levels. Firstly, it ensures the provision of uniform terminology and a structured set of risks, which supports consistency and comparability of analyses. Secondly, it ensures the direct use of external information flows as input data in the risk management process, in accordance with the principle of using the best available information. Thirdly, its modular structure ensures adjustments and extensions without disrupting the basic methodological logic.

It's important to note that the risk register used in this model does not replace the internal risk register; rather, it serves as a complementary layer. External detection of individual risks can trigger reassessment, in-depth analysis, or escalation within existing processes, consistent with the iterative nature of risk management emphasized by ISO 31000.

The risk register enables the extraction of structured, interpretable, and practical signals from heterogeneous, unstructured text data. This model, therefore, exceeds classical approaches based on aggregate indices of uncertainty or sentiment, ensuring multidimensional risk monitoring at the level of individual domains and specific risks. The register presents a necessary condition for a methodologically consistent comparison between GDELT and Nostr and for integrating results into modern risk management systems in non-financial companies.

Methodology

The methodological framework of this article is based on a repeatable analytical procedure that generates quantitative risk activity indicators from text sources. The key requirement of this framework is that it enables comparability between risks and sources (GDELT and Nostr) over time, while maintaining the interpretability of results in the context of corporate risk management. Therefore, the methodology consists of sequential steps: data collection, construction of news volume, standardization, calculation of dynamic indicators, aggregation, and comparison of results.

Let $r_1, r_2, \ldots, r_K$ be all the risks in the risk register ($K = 264$) and let $t \in \{1, \ldots, T\}$ denote a month of the analyzed period. For each data source $s \in \{\mathrm{GDELT}, \mathrm{Nostr}\}$, each risk r_k and each month t we observe a monthly time series of news volume

$$V_{k,t}^{(s)}, k = 1, \ldots, K, \qquad t = 1, \ldots, T.$$

In our analysis $V_{k,t}^{(Nostr)}$ is an absolute number of detections of risk r_k in a month t in social media posts, while $V_{k,t}^{(GDELT)}$ denotes a normalized intensity of news volume of risk r_k in a month t. Although because of gradual development, technical characteristics, and statistical properties, the scales differ, the time series for each risk within each source are consistent, which enables the use of relative and standardized indicators.

The basic indicator of risk exposure is the monthly volume of detections

$$V_{k,t} = \sum_{i=1}^{N_t} \mathbf{1}(d_{t_i} \to r_k),$$

where d_{t_i} denotes individual textual unit (news or a social media post) in a month t and $\mathbf{1}(\cdot)$ is an indicator function, that is equal to 1, if the document d_{t_i} is assigned to risk r_k based on a semantic mapping obtained from the risk register. This indicator measures the intensity of discussions or reporting of individual risks. However, it is not suitable for assessing topicality on its own, as it is heavily dependent on the structural frequency of risk in discourse.

To eliminate bias arising from different base rates of risk occurrence, we standardize the volumes based on the individual risk's history of occurrence. For each risk r_k, month t, and source s, we compute the moving average and the standard deviation over a window of W months (with $W = 12$ in our empirical application):

$$\mu_{k,t}^{(s)} = \frac{1}{W} \sum_{j=0}^{W-1} V_{k,t-j}^{(s)},$$

$$\sigma_{k,t}^{(s)} = \sqrt{\frac{1}{W-1} \sum_{j=0}^{W-1} \left(V_{k,t-j}^{(s)} \mu_{k,t}^{(s)} \right)^2}.$$

The standardized relevance indicator (z-score) is then defined as

$$Z_{k,t}^{(s)} = \frac{V_{k,t}^{(s)} - \mu_{k,t}^{(s)}}{\sigma_{k,t}^{(s)}}.$$

The indicator $Z_{k,t}^{(s)}$ measures how many standard deviations the current reading deviates from the 'normal' level for a given risk. This indicator enables direct comparability between sources, regardless of the absolute scale of volumes. Because risk management often requires the detection of not only high, but also quickly emerging risks, the methodology includes indicators of dynamics or acceleration. The simplest indicator of acceleration is a monthly change of z-score:

$$\Delta Z_{k,t}^{(s)} = Z_{k,t}^{(s)} - Z_{k,t-1}^{(s)}.$$

Positive values of $\Delta Z_{k,t}^{(s)}$ indicate risks that are becoming more relevant, even if the absolute level isn't high. Alternatively, we can use the difference between the short-term and long-term z-score, which enables better trend detection. In empirical results, these indicators are used mostly to identify "emerging risks".

Because the risk register is structured into 22 risk groups, it's strategically better to aggregate indicators for interpreting results. Let G_M be a set of risks from a risk group M. The aggregated indicator of group actuality is defined as an average of z-scores of all risks from the risk group M:

$$Z_{M,t}^{(s)} = \frac{1}{|G_M|} \sum_{r_k \in G_M} Z_{k,t}^{(s)}.$$

Because each risk group has the same number of risks ($midG_M mid = 12$), the aggregation does not require additional weighting and allows for direct comparison between groups. This feature is a direct result of the risk register's balanced design.

To quantify activity concentration between risks, we use the Gini coefficient. Let S_k present all activity of risk r_k over the entire timeframe:

$$S_k = \sum_{t=1}^{T} V_{k,t}.$$

The Gini coefficient is further defined as:

$$G = \frac{\sum_{i=1}^{K} \sum_{j=1}^{K} |S_i - S_j|}{2K \sum_{i=1}^{K} S_i},$$

where K is a number of all risks in the risk register ($K = 264$).

Values close to 1 suggest a high concentration of activity in a small number of risks, which is common in textual data and has important implications for interpreting news volumes.

To better understand the interdependencies among risks, we perform a correlation analysis of time-series of risk z-scores. The correlation coefficient between risks r_i and r_j over the months $t \in \{1, \ldots, T\}$ is defined as:

$$\rho_{\mathrm{ij}} = \mathrm{corr}\left(\{Z_{i,t}\}_{t=1}^{T}, \{Z_{j,t}\}_{t=1}^{T}\right).$$

The correlation matrices enable us to identify risk groups that systematically appear together, which is crucial for understanding systemic risks and developing scenarios.

The key methodological contribution of this article is a comparison of two signal sources. For each month t and risk r_k we calculate the indicators $Z_{k,t}^{(\mathrm{GDELT})}$ and $Z_{k,t}^{(\mathrm{Nostr})}$. The comparison between them enables us to identify four types of situations: consensus of both sources, early signal in Nostr, signal limited to news flow, and absence of signal in both sources. On an aggregate level, we measure the source consistency with z-score correlation and identify top risks by actuality in both sources. The methodology enables the transition from raw text perceptions to standardized, comparable, and interpretable indicators of relevance that can be directly applied in risk management processes.

Results

The analysis of decentralized social media Nostr and global news database GDELT reveals a distinctly heterogeneous structure of risk detections. Figures 3 and 4 show a classification of risks based on the volume of detections in the last month. It's evident that a small number of risks generate a large proportion of all activity, which is consistent with the statistical analysis of concentration in the chapter on the database. This distribution confirms that absolute volume alone is not a suitable indicator of topicality but rather measures the intensity of discussion in specific communities.

A more informative standardized indicator, shown in Figs. 5 and 6, reveals the actual level of risk based on its z-score. From a management perspective, the z-score can be interpreted as an indicator of risks where the observed activity is statistically unusual relative to its own history, regardless of its absolute volume. The highest-ranked risks tend to be topics related to technological disruptions, cyber incidents, and geopolitical tensions, indicating that Nostr acts as a sensitive sensor for rapidly changing topics in the information sphere.

The standardized indicator for GDELT, shown in Fig. 6, highlights risks that have attracted unusually high levels of media attention each month. These risks are often associated with specific events, such as regulatory interventions, armed conflicts, or natural disasters, confirming GDELT's interpretation as a source that responds more strongly to clearly defined events.

The heatmap of z-scores for a wider selection of risks, shown in Fig. 7, provides a comprehensive overview of current dynamics over time. This visualization clearly shows that outbreaks of current events at Nostr are often concentrated in time, and individual risks occur in waves, which is consistent with the high volatility identified in the data analysis.

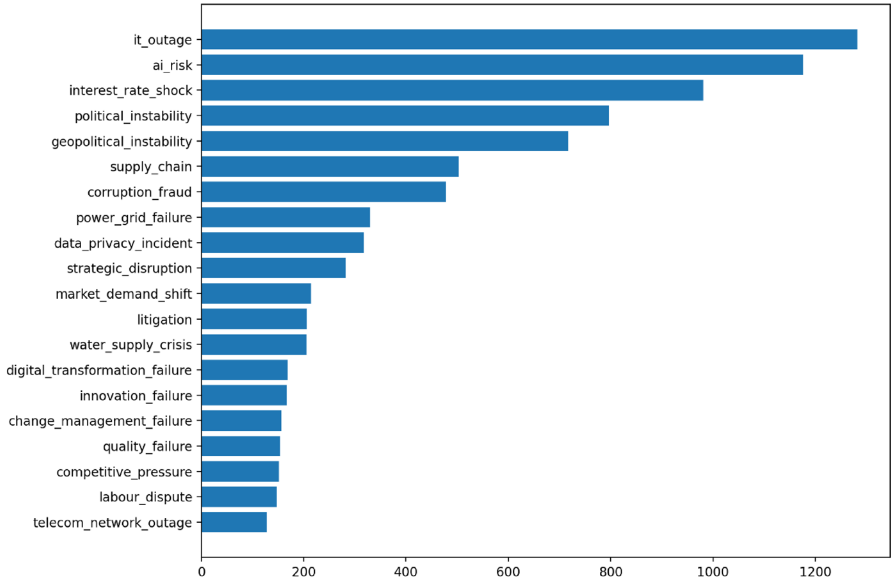

Fig. 3 Risk classification based on absolute monthly volume of detections in the last observed month in the Nostr database. *Source* Author's own calculations. *Note*: The x-axis shows monthly volume, and the y-axis shows the top risks

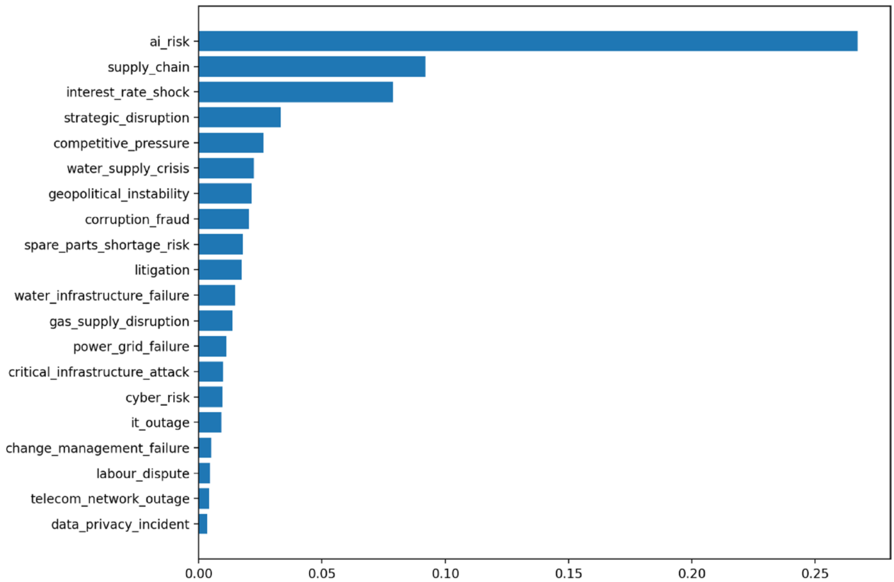

Fig. 4 Risk classification based on absolute monthly volume of detections in the last observed month in GDELT database. *Source*: Author's own calculations. *Note*: The x-axis shows monthly volume, and the y-axis shows the top risks

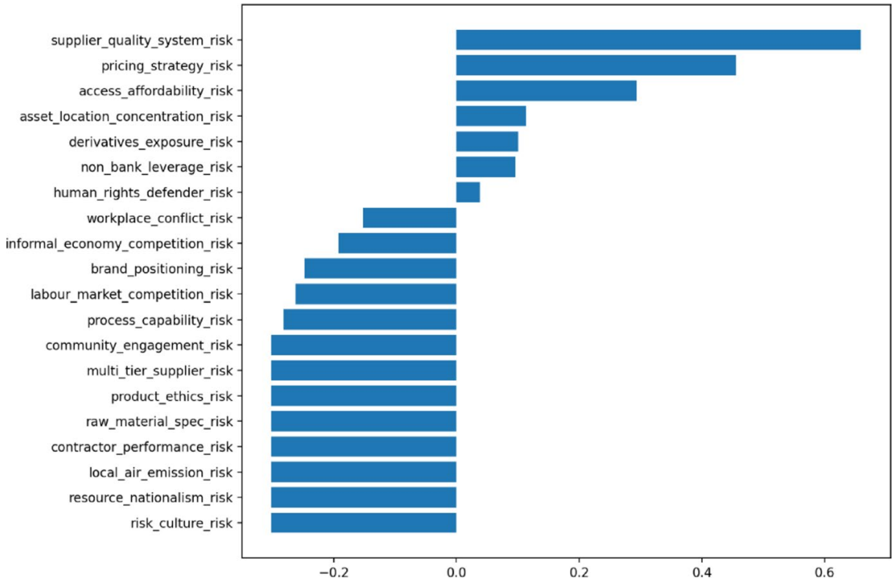

Fig. 5 Distribution of z-score by individual risks in Nostr database. *Source*: Author's own calculations. *Note*: The x-axis shows z-scores, and the y-axis shows the top risks

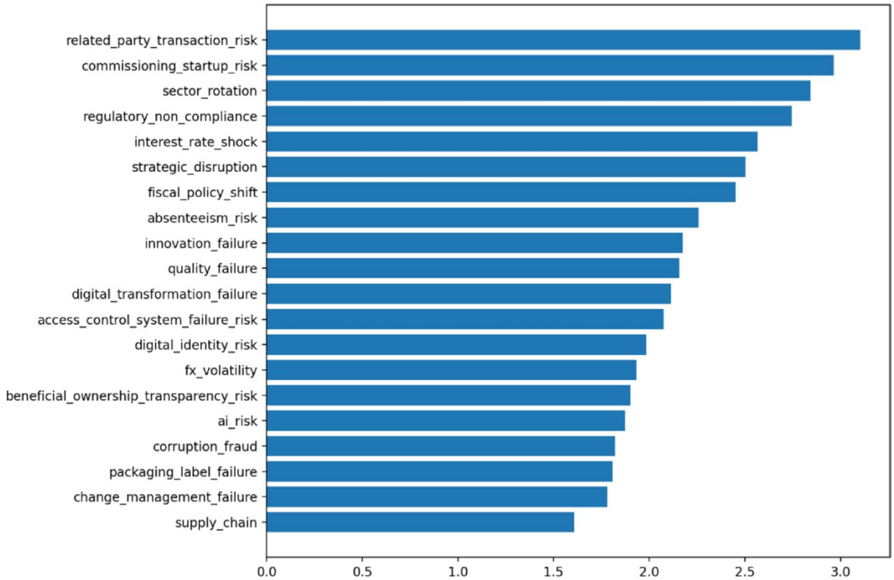

Fig. 6 Distribution of z-score by individual risks in GDELT database. *Source*: Author's own calculations. *Note*: The x-axis shows z-scores, and the y-axis shows the top risks

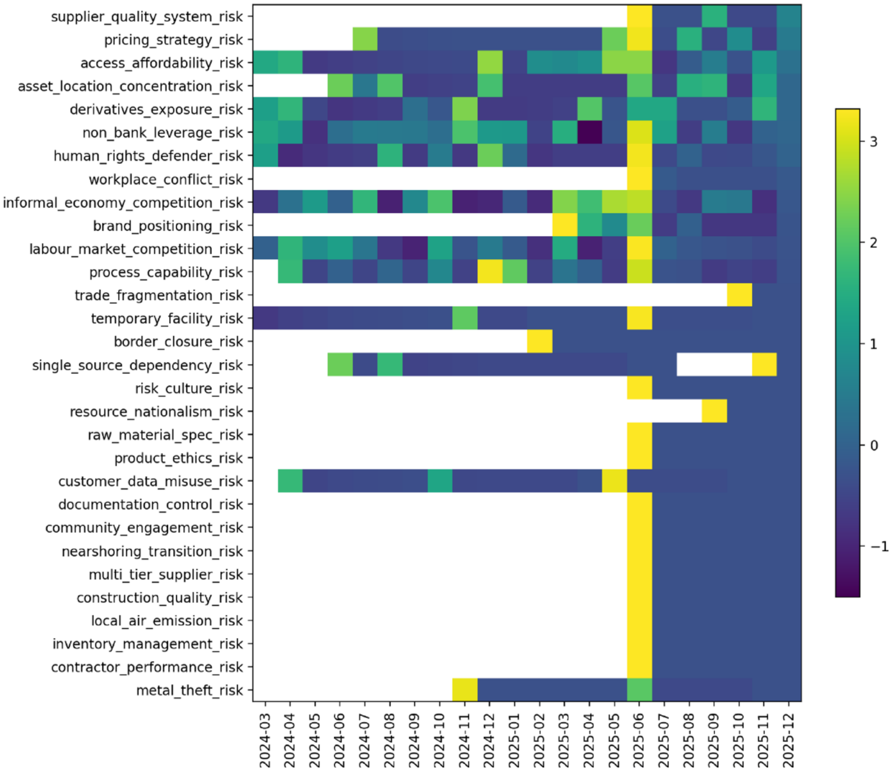

Fig. 7 Heatmap of z-scores (Top 30 Risks). *Source* Author's own calculations

Figure 8 shows the scatter plot of z-score values for Nostr and GDELT in the same month. The quadrant distribution allows for clear interpretation: risks in the upper right quadrant indicate consistency across both sources, while risks that stand out in only one source signal either an early warning or a source-specific discourse.

For company management, discrepancies between news and decentralized signals indicate areas that require additional attention and internal discussion, even in the absence of extensive mainstream media reporting.

The level of consistency between sources is additionally quantified by analyzing the overlap of the top 15 risks by z-score. The discrepancy confirms that Nostr and GDELT capture different aspects of risk perception, which does not necessarily imply a contradiction but rather complements the information. Structural differences between sources are further illuminated by comparing the proportions of risk groups among the top 50 risks, as shown in Fig. 9.

Lastly, Fig. 10 presents the time dynamics of risks that are among the most reported in both sources. This group of risks represents the core of robust signals that are source independent and are therefore particularly relevant for strategic monitoring.

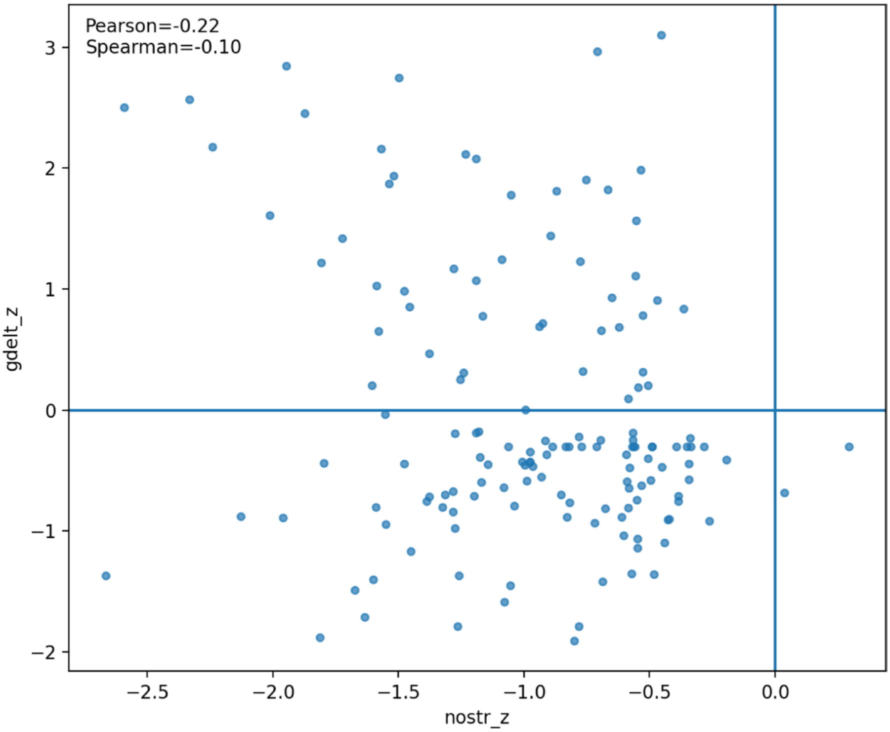

Fig. 8 Scatter Plot of z-score values for Nostr and GDELT for December 2025. *Source* Author's own calculations

Informational value of risk indicators for forward-looking risk assessment

We demonstrated, that the proposed framework is able to detect and structure emerging risks from heterogeneous textual sources. But do these signals contain economically meaningful information beyond descriptive monitoring? We want to determine whether these signals can predict changes in the financial environment, rather than merely reflecting the current situation. To address this, we extend the analysis by linking the constructed risk indicators to forward-looking measures of systematic risk, specifically industry beta coefficients (Jagrič et al. 2026).

We assume that signals are first reflected in the information space, and only then does their impact become apparent in the financial environment. If this assumption is correct, indicators based on text data extracted from news articles and decentralized social networks will need to provide predictive information about future changes in systemic risk. This perspective is consistent with the broader literature on uncertainty indices but extends it by moving from aggregate measures to a structured, multidimensional risk register aligned with corporate risk management frameworks.

To test our hypothesis, we developed a model to predict future industry beta coefficients. We made the prediction based on the beta coefficient's own persistence, macroeconomic conditions, and risk indicators derived from the GDELT database (we

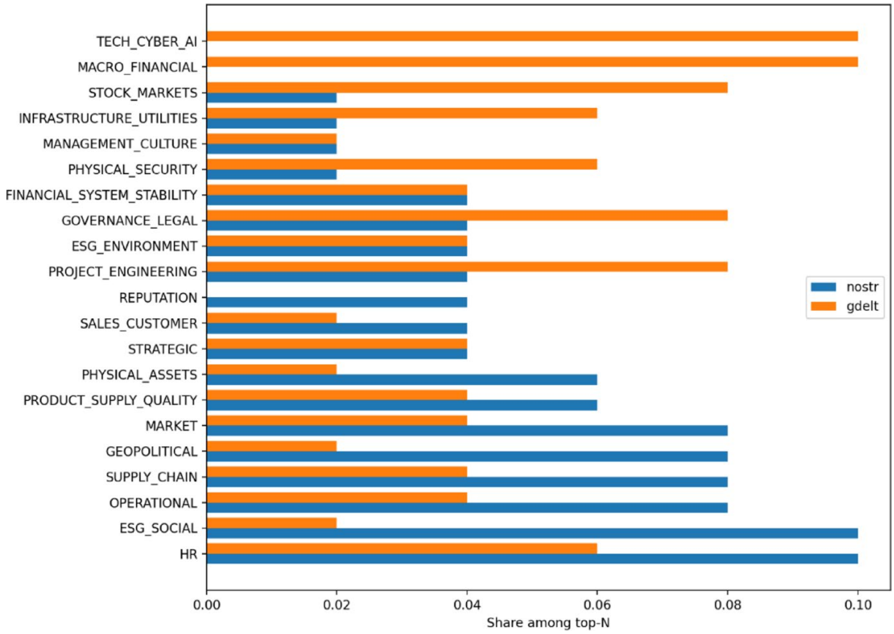

Fig. 9 Comparison of proportions by risk groups of the Top 50 risks (December 2025). *Source* Author's own calculations

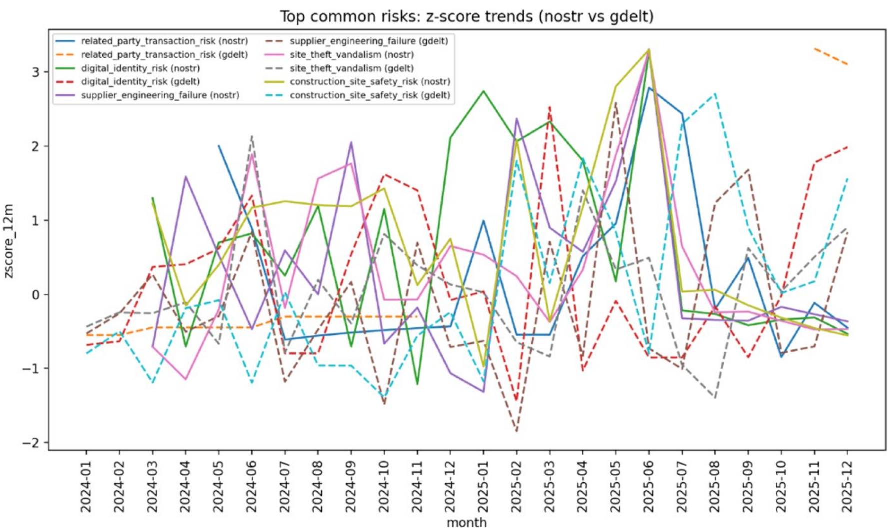

Fig. 10 Time dynamics of relevant risks in both sources. *Source* Author's own calculations

Table 2 Incremental contribution of information sets to beta prediction

Model specification	Avg. explanatory power (R^2)	Relative improvement (%)	Share of predictive contribution (%)
Lagged beta only	0.42	–	100 (baseline persistence)
Lagged beta + macro variables	0.48	+14	62 (beta)/38 (macro)
Lagged beta + risk indicators	0.59	+40	55 (beta)/45 (risk)
Lagged beta + macro + risk indicators	0.68	+62	41 (beta)/23 (macro)/36 (risk)

Source: Author's calculations

chose GDELT because of its larger dataset). All explanatory variables are lagged to ensure strict ex ante interpretation. Risk indicators in the model are expressed as z-scores at the risk-group level. This approach allows for a better understanding of the results while reducing noise in high-frequency data. For each industry i and each month t we defined the econometric model as:

$$\beta_{i,t} = \alpha_i + \gamma_i \beta_{i,t-12} + \boldsymbol{\delta_i M_{t-12}} + \boldsymbol{\epsilon_i R_{t-12}} + \varepsilon_{i,t},$$

where $\beta_{i,t}$ is the industry beta in a month t, α_i is a constant term, $\beta_{i,t-12}$ represents the lagged industry beta (12-month lag), $\boldsymbol{M_{t-12}}$ denotes the vector of lagged macroeconomic variables, $\boldsymbol{R_{t-12}}$ represents a vector of lagged textual risk indicators, and $\varepsilon_{i,t}$ is the error term.

The empirical analysis reveals that the inclusion of risk indicators systematically improves the ability to explain and predict variation in industry betas. Models based solely on lagged beta coefficients capture the persistence of systematic risk in the industry, but they do not account for structural changes in the environment. The inclusion of macroeconomic variables slightly improves predictive power, as they allow us to incorporate the impact of the economic cycle into the model. However, the most noticeable improvement occurs when risk indicators are included, suggesting that external information flows contain predictive signals that are not captured by traditional variables.

The contribution of risk indicators is not uniform across all risk dimensions. Indicators related to supply chain disruptions, geopolitical tensions, technological uncertainty, and changes in market structure appear most frequently as relevant predictors. This finding is consistent with the interpretation that modern systematic risk is increasingly shaped by non-financial factors that materialize through complex global interdependencies. It is important to note that macroeconomic variables only partially capture the risks in the environment. Therefore, supplementing the model with risk indicators is crucial.

To summarize the contribution of different information sets, Table 2 presents a comparative overview of model's performance and variance attribution across alternative specifications. The results are reported as averages across industries and are based on standardized evaluation metrics to ensure comparability.

The results presented in Table 2 highlight several important findings. First, although lagged beta remains a significant predictor due to its persistence, it explains less than half of the total variation when macro and risk indicators are also included in the model. Second, macroeconomic variables provide additional explanatory power, but their contribution is fairly limited given that the model also includes risk indicators. Third, risk indicators derived from text sources make a significant contribution to predictive power, both when used with macroeconomic variables and when used independently.

The difference in predictive power between macroeconomic and risk indicators is particularly noticeable in the specification of the full model. This supports the hypothesis that risk indicators capture information about future changes in the risk environment. Thus, macroeconomic variables reflect current conditions, while risk indicators reflect expectations and early signals.

From a methodological perspective, this extension addresses a key limitation of many text-based approaches, which are often criticized for lacking clear economic validation. By demonstrating that risk indicators improve the prediction of future systematic risk, we establish a direct link between information dynamics and financially relevant outcomes. This strengthens the interpretation of the proposed framework as more than a descriptive monitoring tool, which justifies its importance for risk management systems.

From a practical standpoint, the results imply that risk monitoring systems can be extended beyond static dashboards toward forward-looking analytical tools. Early signals detected in decentralized sources are typically more volatile, but they also identify current trends more quickly. Therefore, they can be viewed as leading indicators of systemic risk. More stable signals, on the other hand, are detected from global news sources. Therefore, for the most effective risk monitoring, it is recommended to use a combination of both sources, which enables a richer understanding of risk dynamics.

Overall, the integration of risk detection and forward-looking validation shows that the proposed framework identifies information that is genuinely useful for economic and financial decision-making. In the model, we linked high-frequency text signals to systematic risk, thereby establishing a connection between qualitative information flows and quantitative financial analysis. This is of considerable importance both for theory and for risk management practice.

Conclusion

In this article, we developed and empirically demonstrated a model for detecting current risks in the environment of non-financial companies, based on a systematic analysis of the external information environment using two freely accessible but conceptually different data sources: a global news database, GDELT, and a decentralized social media platform, Nostr. The key finding of this research is that risk detection is a multidimensional process: no single source of information is sufficient to provide a complete picture of risks; rather, a combination of complementary signals is crucial. The suggested model does not replace traditional risk management practices, but rather complements them by providing a continuous, data-driven contribution to management and supervisory boards' decision-making.

Empirical data clearly show that the structure and dynamics of detected risks differ between sources. GDELT ensures a relatively stable and inertial signal that is strongly linked to specific events and editorially filtered content and therefore reflects a consolidated news perception of risks. Nostr works as a highly volatile and impulsive source, sensitive to early changes in discourse, niche topics, and the rapid spread of discussion in decentralized communities. A comparative analysis using standardized indicators of topicality and dynamics reveals that the degree of direct overlap between sources is relatively low, which does not imply a methodological weakness but rather confirms their conceptual complementarity.

The key contribution of the article is the conceptual and methodological integration of heterogeneous signals into a single risk register, designed in accordance with established risk management frameworks, in particular COSO ERM and ISO 31000. With the use of balanced taxonomy, which evenly covers risk content domains, the model enables direct mapping of external information in a structure that is understandable and usable for risk management. The standardization of volumes based on the history of individual risks further ensures that the indicators measure not only absolute attention but also statistically significant deviations, which are essential for detecting current and emerging risks.

From a research perspective, this article adds to the literature on textual risk detection in several ways. Previous studies have mainly focused on aggregate uncertainty indices or single information sources, while our framework uses two information sources and structured risk register. By integrating signals from both global news and decentralized social media sources, the model captures richer risk signals than the existing models. Therefore, the model contributes to the development of data-driven risk management approaches by demonstrating how large-scale textual information can be systematically incorporated into risk monitoring frameworks.

From a risk-management perspective in companies, the data shows important implications. Firstly, the model enables supplementing classic, predominantly qualitative risk assessments with empirically supported, high-frequency indicators from the external environment. Secondly, the comparative use of multiple signal sources allows for differentiation between robust risks, where there is consensus in the information space, and risks that appear only in certain segments of discourse and therefore require additional analytical judgment. Thirdly, aggregating the results at the risk group level provides a strategic view that is particularly suitable for use at the level of management and supervisory boards, where an overview of broader trends is needed, rather than just individual operational risks.

The empirical results suggest that high-frequency signals from news provide an economically relevant perspective on current risks. Empirical results indicate that these indicators improve the prediction of future systematic risk, such as industry betas, even when accounting for macroeconomic conditions and lagged risk measures. This suggests that information extracted from news flows contains forward-looking signals about changes in the risk environment that are not fully captured by traditional financial or macroeconomic indicators. As a result, news-based indicators can serve as a valuable forward-looking input to corporate risk management, supporting earlier identification of potential risks and improving strategic decision-making.

On the level of economic policy and regulatory supervision, the developed model acts as a complementary tool for monitoring systemic risks. Like existing indices of economic, political, and geopolitical uncertainty, the multidimensional risk-detection model can serve as an early warning mechanism, alerting to shifts in risk perception before they materialize in macroeconomic indicators. Relevant is the ability to monitor differences between news and decentralized social signals, as these discrepancies may indicate emerging topics or a gap between institutional and non-institutional discourse.

Despite these advantages, the approach presented also has significant limitations. The model measures detected and communicated risks in the information space, rather than the direct probability or economic impact of individual events. The results are therefore highly dependent on the structure and bias of the sources used, including editorial practices in news and the composition of user communities in decentralized networks. Our current implementation also relies on a predefined risk register and keyword-based semantic mapping, which may not fully capture all contextual nuances in textual data. In addition, a relatively short time horizon of data is available for decentralized social media Nostr, which limits the ability to evaluate long-term risk trends. Further research should therefore include validating indicators against actual events, financial losses, or operational disruptions, and explore the possibility of weighing individual sources based on the context of use.

Future upgrades to the model may also include expanding semantic processing with more advanced language models, introducing a multilingual risk register, and integrating internal company data with external signals. Despite these open questions, this article's results show that the systematic and standardized analysis of external data sources is feasible and methodologically sound, and represents an important step towards more data-driven and proactive risk management in modern companies.

Acknowledgements Portions of the manuscript were translated from Slovenian to English using the AI-based tool DeepL Translator (DeepL SE, https://www.deepl.com). The authors verified the accuracy and appropriateness of all translations. Furthermore, the manuscript was subjected to linguistic refinement using Grammarly. All suggested revisions were carefully reviewed and approved by the authors.

Author contributions T.J. developed the model and prepared the figures. Both authors wrote the main manuscript, designed the methodology, and reviewed the final version of the manuscript.

Funding Funding was provided by the Univerza v Mariboru.

Data availability The study is based on publicly available data from the GDELT Project and publicly accessible data retrieved from the decentralized Nostr protocol. Documentations are avaliable at https://www.gdeltproject.org/data.html (GDELT) and https://github.com/nostr-protocol/nips (Nostr).

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Baker, S. R., N. Bloom, and S. J. Davis. 2016. Measuring economic policy uncertainty. *The Quarterly Journal of Economics* 131 (4): 1593–1636. https://doi.org/10.1093/qje/qjw024.

Barakat, A., S. Ashby, P. Fenn, and C. Bryce. 2018. Operational risk and reputation in financial institutions: Does media tone make a difference? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3042727.

Caldara, D., and M. Iacoviello. 2022. Measuring geopolitical risk. *American Economic Review* 112 (4): 1194–1225. https://doi.org/10.1257/aer.20191823.

Chen, P., and M. Ji. 2025. Deep learning-based financial risk early warning model for listed companies: A multi-dimensional analysis approach. *Expert Systems with Applications* 283 : 127746. https://doi.org/10.1016/j.eswa.2025.127746.

Committee of Sponsoring Organizations of the Treadway Commission. 2017. *Enterprise risk management: Integrating with strategy and performance*. COSO. https://www.coso.org.

Groth, S. S., and J. Muntermann. 2011. An intraday market risk management approach based on textual analysis. *Decision Support Systems* 50 (4): 680–691. https://doi.org/10.1016/j.dss.2010.08.019.

Hassan, T. A., S. Hollander, L. van Lent, and A. Tahoun. 2019. Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics* 134 (4): 2135–2202. https://doi.org/10.1093/qje/qjz021.

Hong, D., Z. Fu, X. Zhang, and Y. Pan. 2025. Research on the development and application of the GDELT event database. *Data* 10 (10) : 158. https://doi.org/10.3390/data10100158.

Hopp, F. R., J. Schaffer, J. T. Fisher, and R. Weber. 2019. iCoRe: The GDELT interface for the advancement of communication research. *Computational Communication Research* 1 (1): 13–44. https://doi.org/10.5117/CCR2019.1.002.HOPP.

Huang, H., and T. S. Lim. 2024. Construction and optimization of financial risk management model based on financial data and text data influencing information system. *Journal of Information Systems Engineering and Management* 9 (2) : 24534.

Jagrič, T., S. O. Grbenic, M. Galun, and A. Herman. 2025. Identifying key risks for nonfinancial companies. *Global Risk Radar* 1 (1): 3–12.

Jagrič, T., S. Grbenic, and A. Herman. 2026. Forecasting industrial beta one year in advance: Macroeconomic and artificial intelligence-supported risk factors in the US economy. *International Review of Economics and Finance* (under review).

Ma, W., K. Li, W. F. Pan, and X. Wang. 2025. Measuring systemic risk in China: A textual analysis. *China Finance Review International* 15 (4): 669–687. https://doi.org/10.1108/CFRI-06-2023-0155.

Nostr Protocol Developers. 2023. *Nostr protocol specification (NIPs)*. https://github.com/nostr-protocol/nips.

Pei, J., S. Vadlamannati, L.K. Huang, D. Preotiuc-Pietro, and X. Hua. 2025. Modeling and detecting company risks from news: A case study in Bloomberg News. *arXiv*. https://arxiv.org/abs/2508.10927.

Raman, A., S. Joglekar, E. De Cristofaro, N. Sastry, and G. Tyson. 2019. Challenges in the decentralised web: The Mastodon case. In: *Proceedings of the ACM Internet Measurement Conference (IMC '19)*: 217–229. https://doi.org/10.1145/3355369.3355572.

International Organization for Standardization. 2018. *ISO 31000:2018: Risk management—guidelines*. ISO. https://www.iso.org/standard/65694.html.

Roeder, J., M. Palmer, and J. Muntermann. 2022. Data-driven decision-making in credit risk management: The information value of analyst reports. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2022.113770.

Stander, Y. S. 2024. A news sentiment index to inform International Financial Reporting Standard 9 impairments. *Journal of Risk and Financial Management* 17 (7) : 282. https://doi.org/10.3390/jrfm17070282.

The GDELT Project. 2022. *GDELT 2.1 Global database of events, language, and tone*. https://www.gdeltproject.org/data.html.

Wei, Y., and G. Tyson. 2025. An empirical analysis of the Nostr social network: Decentralization, availability, and replication overhead. *Proceedings of the ACM on NetworkIng* 3:1. https://doi.org/10.1145/3768994.