

Semantic Similarity Measures in Newspaper Text for Detecting and Predicting Disruptive Institutional Events

Laura Mayoral* Hannes Mueller† Margherita Philipp‡
Christopher Rauh§ Renato Vassallo¶

January 14th 2026

Abstract

This article proposes a semantic-similarity approach to detecting and predicting rare events in newspaper text, and applies it to institutional disruptions. Using a global news corpus covering more than 170 countries, we measure the similarity of headlines to event-specific prototypes in embedding space and aggregate these signals to identify disruptions to political institutions. We combine these text-based measures with supervised nowcasting and targeted human verification to expand existing datasets on military coups, irregular term-limit extensions, and weakening of the judiciary. The resulting event data are then used to forecast the likelihood of disruptions up to 12 months ahead, providing a high-frequency and scalable tool for monitoring institutional risk. As an illustration of its empirical value, we document that coups are followed by large and persistent declines in economic growth. More broadly, the framework can be adapted to detect and track a wide range of economic and political events and policy actions from news text in real time and in historical archives.

Keywords: Political Institutions, Autocratization, Military Coups, Term Limit Evasion, Judiciary Weakening, Semantic Similarity, Embeddings, Nowcasting, Forecasting.

JEL Codes: C53, C55, D72, P16

*Institut d'Anàlisi Econòmica (IAE-CSIC), BSE, and CEPR. mayoralaura@gmail.com.

†IAE-CSIC, BSE, CEPR. h.mueller.uni@gmail.com

‡Fundació d'Economia Analítica (FEA). margherita.philipp@bse.eu

§IAE-CSIC, BSE, University of Cambridge, CEPR. cr542@cam.ac.uk

¶Universitat Autònoma de Barcelona, BSE. renato.vassallo@bse.eu

‖We acknowledge support from the German Federal Foreign Office. Mayoral, Mueller & Rauh acknowledge support from the Severo Ochoa Programme for Centres of Excellence in R&D (BSE CEX2024-001476-S), funded by MCIN/AEI/10.13039/501100011033. Mayoral & Mueller acknowledge support from grant PID2024-160959OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF EU, & the AGAUR-Generalitat de Catalunya (2021-SGR-416); Rauh acknowledges support from the ATRAE grant AEI/MICINN (ATR2023-144291).

1 Introduction

Political institutions around the world are under pressure amid a worrying trend toward autocratization (Razin 2024). Sub-Saharan Africa has experienced a surge in military coups, with nations such as Mali and Guinea succumbing to sudden power grabs. Simultaneously, established democracies face mounting challenges as populist leaders erode checks and balances, jeopardizing the integrity of democratic institutions. These events do not only have political consequences at the national and international level, but may also have negative economic implications (Acemoglu et al., 2019; Boese-Schlosser and Eberhardt, 2025; Funke et al., 2020). However, currently there are no continuously updated databases of related incidents. A reliable and up-to-date registry is required both to study the impact of disruptive events in the past and predict their occurrence in the future.

This article presents a system designed to fill these gaps by proceeding in two steps. First, disruptions of institutions across the globe are detected by leveraging a combination of existing datasets and an extensive news text corpus covering more than 170 countries. The methodology combines existing hand-coded events with zero-shot semantic similarity features extracted from a large news corpus, together with low-label supervised nowcasting and iterative human validation. This approach—combining existing hand-coded events with zero-shot semantic similarity features and human validation—allows us to construct a comprehensive database of past events that disrupted political institutions. Events include military coups, the manipulation of term limits, and attacks on the judiciary. This approach ensures that various forms of autocratization are captured and monitored effectively. Second, using a similar set of features to nowcast disruptive events, a forecasting model is developed to predict the likelihood of disruptive events occurring during the next year.

Our contribution is threefold. First, we introduce a semantic-similarity framework for event detection in newspaper text that does not rely on hand-crafted dictionaries and is therefore well suited to cross-country settings where terminology and reporting styles vary substantially. Second, we combine these similarity signals with supervised nowcasting and targeted human validation to expand existing datasets on institutional disruptions, producing a high-frequency panel with transparent coding rules and traceable text evidence. Third, building on this measurement layer, we develop a forecasting model that generates monthly updates of disruption risk over the subsequent year, providing a practical early-warning tool for researchers and policymakers. More broadly, the proposed architecture can be readily adapted to detect and monitor a wide range of economic and political events and policy actions in both real-time streams and historical news archives.

To detect events in unstructured news text, we use modern text-embedding models that represent each headline as a numerical vector summarizing its meaning into a latent space. Headlines with similar content are placed close to one another in this vector space, even if they use different words or expressions. This provides a substantial advantage over keyword searches, which can miss relevant articles when an event is described indirectly or flagged by unusual terminology. For each event category—coups, term-limit changes, and judicial interference—we generate a small set of representative example headlines and compute their average position in the embedding space.¹ These serve as reference points. New headlines are compared to these references, and those more similar to them are more likely to describe institutional disruptions.

A concrete example illustrates how this works in practice. Consider Tunisia, where the limited ground-truth data code a sequence of judicial weakening events in early 2022, including the dissolution and replacement of the Supreme Judicial Council and subsequent institutional changes that increased executive control over election oversight. In the corresponding news coverage, we observe headlines such as “Tunisia president affirms dissolution of judicial watchdog.” This headline is mapped to the same region of embedding space as our judiciary-interference examples, even though it does not rely on a narrow vocabulary such as “court packing” or “constitutional crisis.” In particular, its semantic content is close to prototype headlines such as “President Expands Control Over Judiciary: New Law Alters Constitutional Court Powers,” which reflects a similar mechanism of executive intervention that undermines judicial independence. As a result, the similarity score increases in months when judicial institutions are being dismantled or restructured, allowing the model to detect this type of institutional disruption even when descriptions differ across countries and reporting styles.

By aggregating these similarity measures across all headlines published in a country during a given month, we obtain clear temporal patterns: months containing institutional disruptions exhibit marked increases in similarity relative to other periods. Because this method captures meaning rather than specific vocabulary, it performs reliably across countries with very different reporting styles. These aggregated similarity measures then serve as inputs to the supervised nowcasting model that determines whether a disruptive event occurred.

¹We focus on coups, term-limit changes, and judicial interference because these represent three of the most consequential and commonly monitored forms of executive overreach in comparative politics. Each involves a clear disruption of institutional checks and balances, is typically covered in domestic and international media, and has well-developed coding rules in existing datasets. These characteristics make them suitable for automated detection in news text and allow meaningful validation against established benchmarks. The proposed method can be extended to other events.

The nowcasting system performs strongly across all three forms of institutional disruption. Using monthly aggregates of the similarity-based text signals, the model successfully distinguishes periods of heightened institutional risk from periods of stability, even for relatively subtle events such as judicial interference. Importantly, its performance compares favorably to traditional dictionary methods: the news-based features capture semantic cues that keyword lists consistently miss, leading to more accurate early warnings and far fewer false alarms. These gains are particularly evident for forms of autocratization that lack standardized language in news reporting, where embedding-based signals offer a clear advantage. This accuracy makes it feasible to retrospectively code institutional disruptions over long historical periods and across a large set of countries, thereby producing an expandable database of events at monthly frequency.

The second step uses these text-based signals to forecast the likelihood of future institutional disruptions. Whereas the first step identifies when events have occurred in the past, the forecasting stage asks whether similar patterns in the news can anticipate emerging risks. To do this, we summarize the information in each country-month, such as the intensity and dispersion of similarity scores, topical shifts in news coverage, and the time elapsed since earlier disruptions, and use these features to predict whether a disruptive event will occur within the following year. Because the underlying data arrive daily but disruptions are rare and irregular, we employ an expanding-window forecasting design that mimics the information set available to a policymaker at each point in time. A random forest model is then trained on all historical data up to a given month, and used to generate a probability of disruption for the twelve months ahead. Repeating this process month by month yields a full pseudo-real-time forecast history, enabling an honest assessment of how well news-based signals anticipate episodes of autocratization. This structure is crucial for operational use: the forecasting model updates as soon as new text becomes available, allowing institutional risks to be monitored with far higher frequency than traditional political indicators.²

The forecasting model delivers meaningful predictive power for all types of institutional disruptions, despite their rarity and heterogeneity. Out-of-sample tests show that months with elevated disruption risk are systematically associated with higher predicted probabilities, and the model performs well not only for coups, where previous forecasting efforts existed, but also for term-limit evasions and judicial interference, for which no comparable real-time prediction systems were available. Although the events are infrequent, the model

²We plan to add monthly updates concerning the likelihood of events disruptive to institutions to the conflict early warning website conflictforecast.org which provides monthly updates of the likelihood and predicted intensity of political violence, as well as riot risk and the risk of sexual violence at the administrative level.

consistently distinguishes high-risk from low-risk periods, with ROC-AUC values around 0.75–0.80 across categories. Precision is modest, reflecting the rarity of these episodes, but improves sharply for “hard onsets”, where disruptions follow long periods of institutional stability. These results show that news-based features contain valuable predictive information and that the approach can provide timely warnings about looming institutional stress in ways that traditional annual governance indicators cannot. Moreover, because of this methodology, forecasts can be updated monthly.³

Our work contributes to a broader tradition in economics that uses text for event detection and policy analysis. A prominent early example is [Romer and Romer \(2004\)](#), who hand-coded narrative sources to identify exogenous monetary policy shocks. More recently, [Juhász et al. \(2025\)](#) apply a carefully constructed dictionary to large historical corpora to systematically identify trade-related events, showing how textual evidence can be leveraged to construct new datasets of economic relevance. We extend this agenda by moving from static event identification to dynamic event detection and forecasting. Specifically, we combine large-scale news corpora, embedding-based similarity measures, and machine-learning methods to detect signals of autocratization and incorporate them into a forecasting system. This enables both the expansion and monthly updating of event datasets and the production of forward-looking risk assessments. In doing so, we link narrative and dictionary-based approaches with automated text-driven methods, providing policymakers with timely information to anticipate and mitigate democratic erosion. The same architecture—semantic prototypes, similarity scoring, and supervised nowcasting—could be readily adapted to detect and predict economic phenomena such as fiscal crises, policy reversals, sanctions episodes, or major regulatory changes.

Finally, we use our constructed database of coups to study their impact on the economy. [Acemoglu et al. \(2019\)](#) show that transitions to democracy lead to sizable and persistent increases in income, highlighting the economic costs of authoritarian rule. We illustrate the flip side of this mechanism by focusing on abrupt institutional breakdowns. We find that coups d’état are followed by sharp and long-lasting declines in GDP per capita, with no evidence of differential trends prior to the event. The magnitude and persistence of these losses mirror the gains associated with democratization documented by [Acemoglu et al. \(2019\)](#), underscoring that institutional reversals are not merely political events but major economic shocks. These findings reinforce the importance of timely detection and monitoring

³Jay Ulfelder provided annual updates for four years until 2015, publishing his predictions on his now dormant [dartthrowingchimp](#) blog post and quarterly updates were provided by One Earth Fund’s (OEF) discontinued CoupCast, created in 2016 with results regularly published on [The Die is Forecast](#) Medium blog post between 2019 and 2021).

of institutional disruptions, motivating the need for high-frequency data and forward-looking tools that can anticipate periods of heightened institutional risk.

2 Data

This study uses two data types to compile a comprehensive list of past institutional disruption events and then forecast them. The first is structured data available from databases, while the second involves text-based features derived from news sources. Using natural language processing (NLP), we extract meaningful insights from this unstructured textual data.

2.1 Structured data sources

Coup d’État: We rely on the Cline Center’s Coup d’État Project, starting in 1950 and updated to the end of 2024 (Peyton et al., 2023), and merge it with data provided by Powell and Thyne (Powell and Thyne, 2011) to expand coverage. While the Powell and Thyne dataset focuses on successful and failed military coups since 1945, the Cline Center dataset also includes conspiracies and categorizes a broader range of non-military coups and coup attempts, such as the storming of Capitol Hill on 6 January 2021.

The combined dataset we work with tracks 1125 coup events for 194 countries, encompassing realized coups, attempts, and conspiracies. To construct a global data panel with monthly frequency from 1989 to the present, we filter for events within this period, resulting in 377 coup-related events.

Term-limit evasion: To expand our target beyond coups d’état, we incorporate term-limit evasions as critical events indicative of autocratization. As noted by the Africa Center for Strategic Studies, “term-limit evasions are linked to higher levels of autocracy, corruption, conflict, and propensity for coups.”⁴

As a starting point, we draw on events identified in the literature. Ginsburg et al. (2011) highlight that, of the 352 cases in their dataset (starting from the early 1900s) where presidents had the opportunity to overstay their constitutionally permitted terms, 89 involved attempts to remain in power, and 71 of those attempts were successful.

Table 1 presents a subsample of events from Versteeg et al. (2020), which provides a comprehensive global survey of term-limit evasion episodes since 2000. These events predominantly pertain to African and Latin American countries.

⁴See <https://africacenter.org/spotlight/term-limit-evasions-coups-africa-same-coin/>.

Table 1: Term-limit evasion events by strategy

Evasion strategy	Description	Country	Year	Result	President
Amendment	Amend the constitution to eliminate term limits or extend number of terms	Cameroon	2008	Successful	Biya
		Senegal	2012	Successful	Wade
		Djibouti	2010	Successful	Guelleh
		Rwanda	2015	Successful	Kagame
		Burundi	2020	Successful	Nkurunziza
		Zambia	2001	Attempt	Chiluba
		Malawi	2002	Attempt	Mulizi
		Nigeria	2006	Attempt	Obasango
		Senegal	2011	Attempt	Wade
		Burkina Faso	2014	Attempt	Campaoré
		Colombia	2010	Attempt	Uribe
Court	Rely on legal theory or the courts to reinterpret term limits	Honduras	2009	Attempt	Zelaya
		Paraguay	2017	Attempt	Cartes
		Burundi	2015	Successful	Nkurunziza
		Nicaragua	2009	Successful	Ortega
Referendum	Use referenda to seek public approval	Honduras	2015	Successful	Hernandez
		Bolivia	2016	Successful	Morales
		Tajikistan	2016	Successful	Rahmon
		Panama	1998	Attempt	Perez
		Bolivia	2016	Attempt	Morales

Note: Subsample of events derived from Versteeg et al. (2020).

In total, we identified 117 events related to our second autocratic category of term limit evasions.

Weakening of the judiciary: Another way in which democracy can be undermined is through weakening of the judiciary by the executive or legislative branches. This strategy either ensures that the courts align with the preferences of the incumbent government or creates an environment where judges who dissent find it nearly impossible to express their genuine views.

A significant branch of the literature has examined episodes of judicial manipulation in Latin American countries. One of the most notable contributions is that of Helmke (2017), who documents instances of judicial manipulation in 18 Latin American countries between 1985 and 2008.

Both the attempts and successful cases of judicial manipulation presented in Table 2 will serve as the foundation for constructing our third autocratic category: *judiciary-weakened*.

Table 2: Presidential attacks on the courts in Latin America, 1985–2008

Country	Administration	Year*
Argentina	Alfonsín	1987
	Duhalde	2002
	Menem	1989
	N. Kirchner	2003
Bolivia	Paz Zamora	1990
	Morales	2006; 2007; 2008
	Sánchez de Lozada	1993
Chile	Aylwin	1991
	Frei	1997
Ecuador	Bucaram	1996
	Cordero	1985
	Durán-Ballén	1994
	Gutiérrez	2004; 2005
	Correa	2007
Guatemala	Serrano	1993
Nicaragua	Bolaños	2004
Paraguay	Wasmosy	1993
	Duarte	2003
Peru	Fujimori	1991; 1997
Venezuela	Pérez	1992
	Chávez	1999; 2002; 2003

Note: This table presents judicial crisis events identified by Helmke (2017), spanning Latin American countries from 1985 to 2008. Bolded years indicate successful attacks.

A further source of data on judicial weakening is the Democratic Erosion Event Dataset (DEED) by [Gottlieb et al. \(2025\)](#). It contains more than 22,000 events related to democratic erosion and autocratic consolidation in 152 countries from 2000 to 2023. They distinguish between events that lead to erosion (precursors) and events where erosion is institutionalized (symptoms). We only include a subset of events in our target, namely the precursor ‘Delegitimizing or Weakening the Judiciary’, and the symptom ‘Reduction in Judicial Independence’. The former is rhetorical, i.e. includes instances of verbal attacks or threats, while the latter refers to institutional changes, such as court packing or reforms that compromise impartiality ([Democratic Erosion Consortium, 2025](#)). We include 149 events from these subcategories.

2.2 Unstructured textual data

Our text corpus consists of over six million documents from 1989 to the present. The news articles are sourced from Factiva and Lexis Nexis. These include 400,000 articles from the New York Times, 100,000 from The Economist, 1 million from the Associated Press, 3.5 million from BBC Monitor, and 10,000 from LatinNews. Articles are downloaded based on queries requiring a country or capital name in the title or lead paragraph. While news data may exhibit biases, particularly under media censorship, [Mueller and Rauh \(2022\)](#) demonstrate that this does not significantly affect conflict prediction models. LatinNews is included to enhance coverage for Latin America, complementing BBC Monitor’s focus on Asia and Africa. The headlines of these newspaper articles are what are used for the semantic similarity scores.

In order to use the newspaper text more broadly at the forecasting stage, we reduce the newspaper articles to topics. Standard NLP preprocessing is applied, including punctuation removal, stop word elimination, and lemmatization. We analyze unigrams, bigrams, and trigrams, excluding tokens that are either too frequent (appear in half of the documents) or too rare (fewer than 200 documents). The resulting corpus organizes documents at the country-month level, providing a comprehensive monthly news landscape for every country from January 1989 to the present. This text data is used to generate predictive features. We use Latent Dirichlet Allocation (LDA) to generate features at the country-month level, linking broad text topics to detected autocratization events. LDA, introduced by [Blei et al. \(2003\)](#), is a probabilistic model that assumes documents are mixtures of latent topics, and topics are mixtures of words.⁵

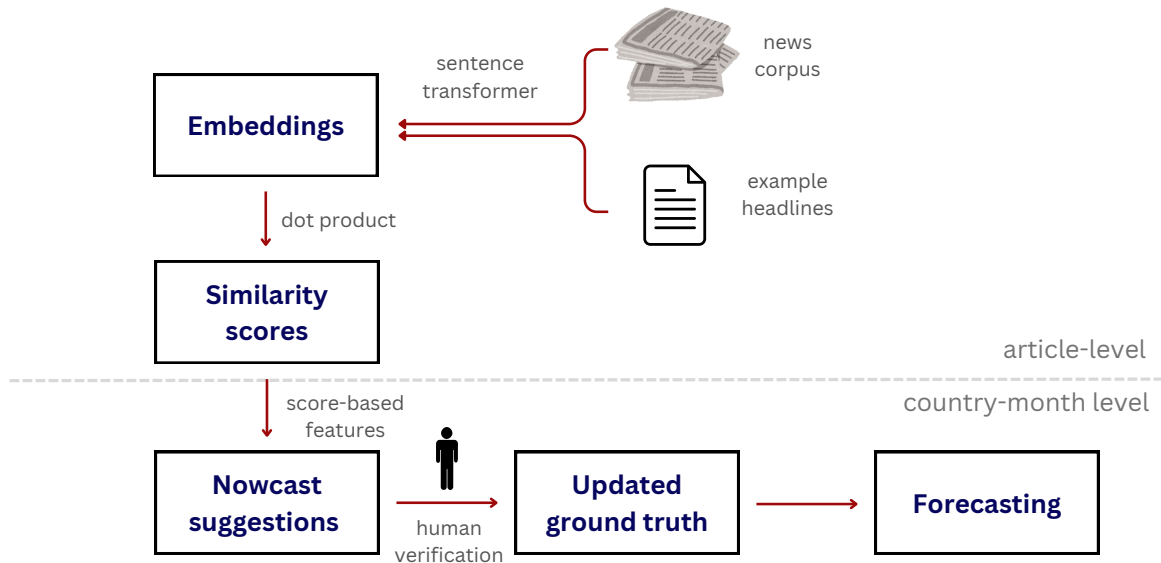
3 Methodology

The architecture utilized in this study is illustrated in [Figure 1](#). It begins with the management of our text corpus, with the article headlines being reduced into embeddings using a sentence transformer. By representing text as numerical vectors in a shared semantic space, we enable meaningful comparisons between headlines based on their meaning rather than exact wording. Semantic similarity methods are then used to score each headline against small sets of representative examples, producing probabilities that a headline relates to one of the disruptive events of interest.

⁵Our implementation relies on the Python package from [Řehřek and Sojka \(2010\)](#) and the dynamic LDA model by [Hoffman et al. \(2010\)](#), estimating 15 topics. This approach reduces text dimensionality, allowing the model to identify useful features for forecasting autocratization events. For further details, see [Mueller et al. \(2024\)](#).

This disaggregated, article-level information feeds into a supervised algorithm for nowcasting events at a monthly frequency, a process further refined through human validation. The updated target dataset, combined with additional features, becomes the input for forecasting algorithms that predict the likelihood of disruptive events up to 12 months ahead.

Figure 1: Overview of the pipeline architecture



Notes: News headlines as well as example headlines for each event type are being embedded and then compared to generate similarity scores. Aggregating from articles to the country-month level allows for the generation of score-based features for the nowcast. A human in the loop checks the suggested country-months for each event type and updates the ground truth for the latest month(s) before running a new forecast.

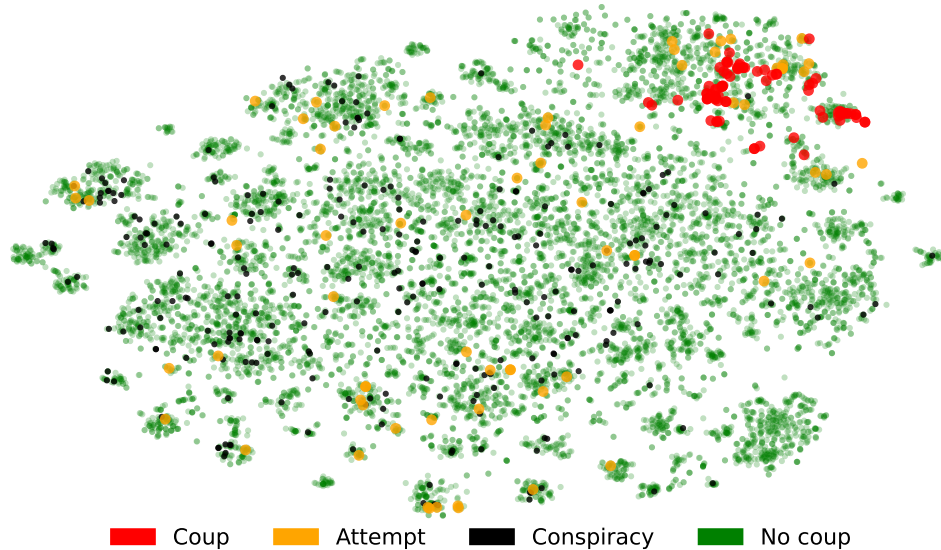
3.1 Reducing dimensionality of text

The process begins by transforming headlines into embeddings—a numerical representation of text—using a sentence embedding model. We use *all-MiniLM-L6-v2*, which was finetuned on the 6 layer version of the already distilled *MiniLM-L12-H384-uncased* model (Wang et al., 2020). *all-MiniLM-L6-v2* demonstrated superior accuracy, and projects textual input into a 384-dimensional vector space.⁶ For illustrative purposes, we reduce the 384 dimensions to two dimensions using t-SNE in Figure 2 and plot newspaper headlines for the Democratic

⁶We explored alternative sentence embedding models. Appendix A.4 provides a detailed comparison of embedding models, showcasing their performance and computational requirements.

Republic of Congo. The red dots mark headlines in a two-week window around successful coups, which clearly are clustered in a similar space.⁷

Figure 2: Two-dimensional representation of article embeddings during coups in the Democratic Republic of the Congo



Notes: We use t-SNE to reduce the 384 dimensional embedding vectors of each headline pertaining to the Democratic Republic of the Congo into a two dimensional space. Red dots represent headlines around successful coups, orange around attempted coups, and black dots around coup conspiracies. The green dots represent the remaining headlines.

3.2 Scoring articles with semantic similarity scores

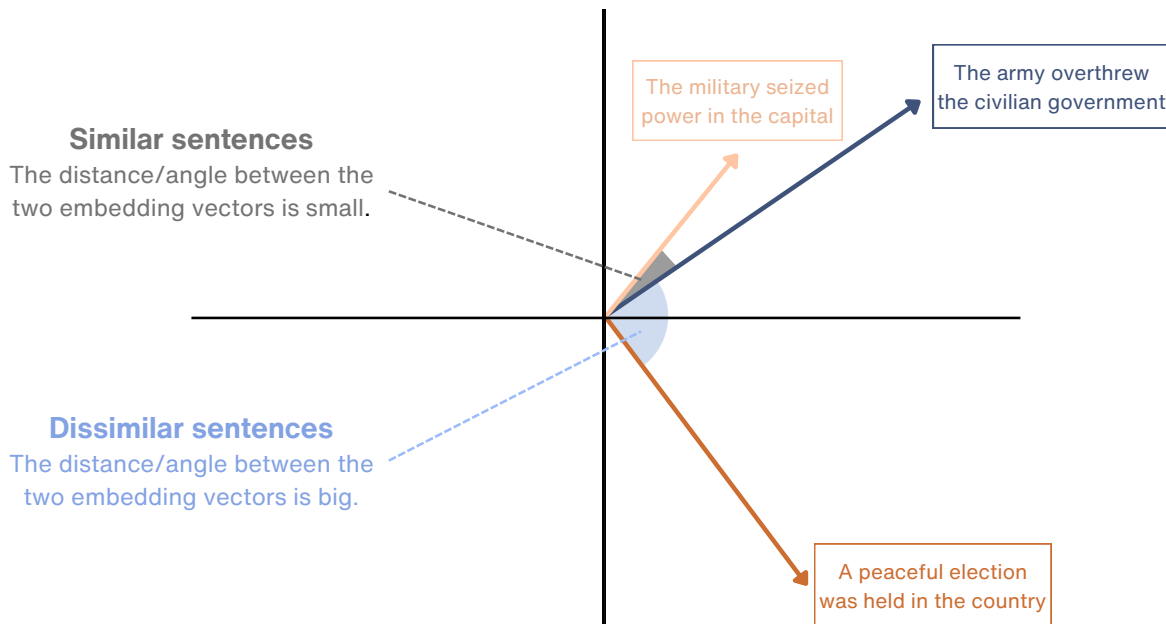
This step assigns semantic similarity scores to news headlines to assess their relevance to different types of institutional disruptions. Using the text-embedding model described above, we represent both real headlines and a small support set of manually curated example headlines—covering coups, term-limit evasions, and judicial interference—in a shared vector space. Curated examples serve as semantic prototypes, and new headlines that are more similar to prototypes of a category receive higher similarity scores. This approach is particularly useful when labeled data are scarce or costly to produce, as it leverages the structure of the embedding space rather than requiring extensive manual annotation.

The support sets consist of 40-70 prototype headlines per category, e.g. ‘Authorities confirm

⁷In Appendix Figure A2 we show the analog representations of headlines in Cameroon for term limit evasion and the Philippines for judiciary weakened.

coup d’etat underway, leaders claim control’ forms part of the prototype sentences for coups (see Appendix A.3 for a full list). The example sentences of each category are converted into embeddings and each news headline in the corpus is compared to each of the prototype vectors. Headlines published around institutional disruptions should be positioned closer to the prototype phrases in embedding space, as illustrated in Figure 3.

Figure 3: Illustrative representation of headlines in the embedding space



Notes: Each vector corresponds to a headline, with proximity indicating similarity. The examples are fictitious.

The similarity or dissimilarity between a headline u and a predefined example v from the support set is measured using cosine similarity, which is computed as

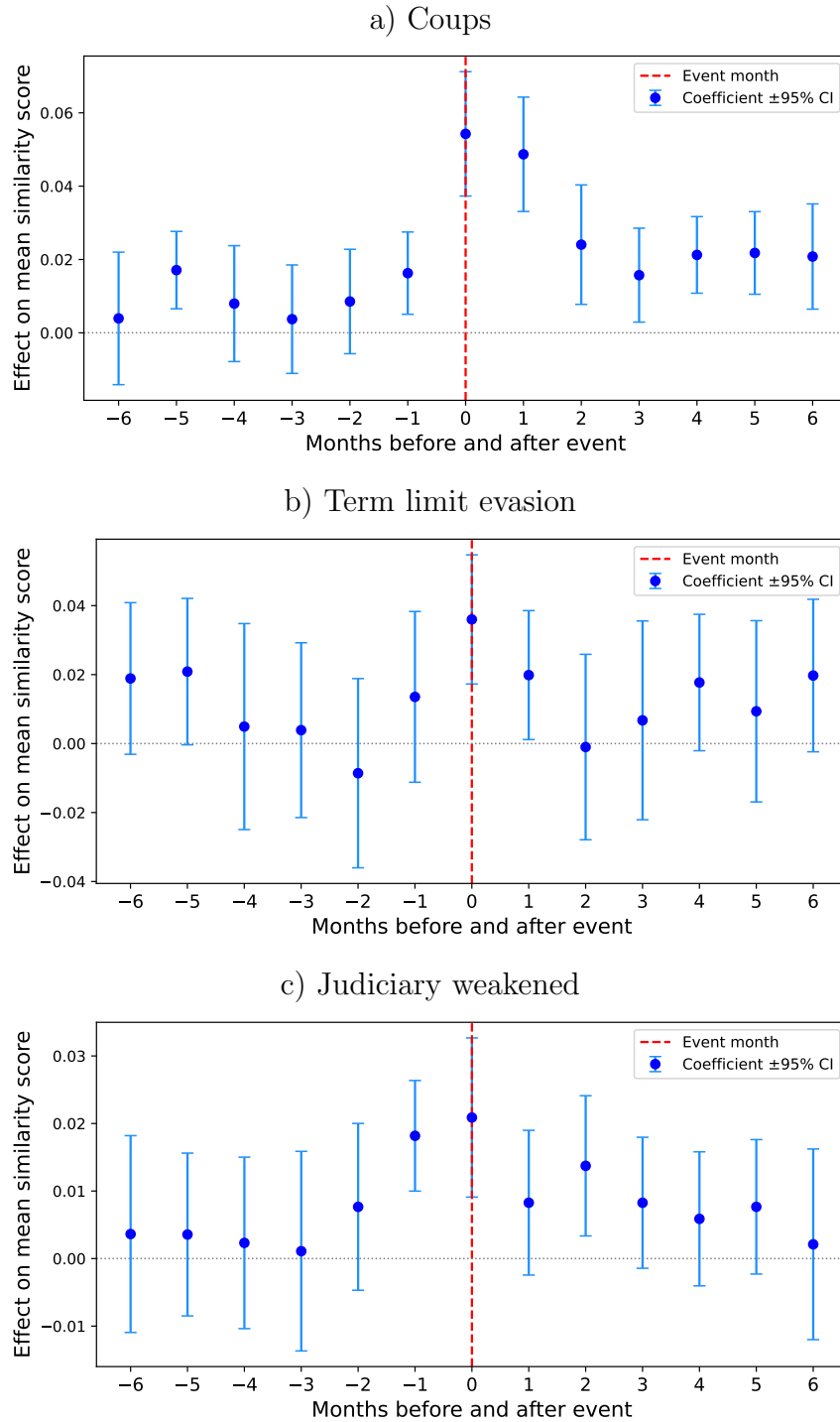
$$\text{cosine similarity}(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}.$$

The distribution of scores around events can then be used downstream to detect and predict whether given events have occurred in the past or will occur in the future. In Figure 4 we show how similarity scores behave around events. In panel a) we see that mean similarity scores increase significantly the month before a coup by almost 0.02, and then shoot up by almost 0.06 in the month of the coup. The score remains significantly elevated for the following six months. For term limit evasion in panel b) we see a different pattern. There appears to be no anticipatory increase in the mean similarity score, but a significant increase

of 0.04 in the month of the event. The month after the term limit evasion, the the mean score drops back to slightly above its mean. In panel c) we see that for a weakened judiciary, there is a significant increase preceding the month of the event of 0.02 and a slightly higher increase in the actual month of the event. In the following months the score reverts back to slightly above its mean.

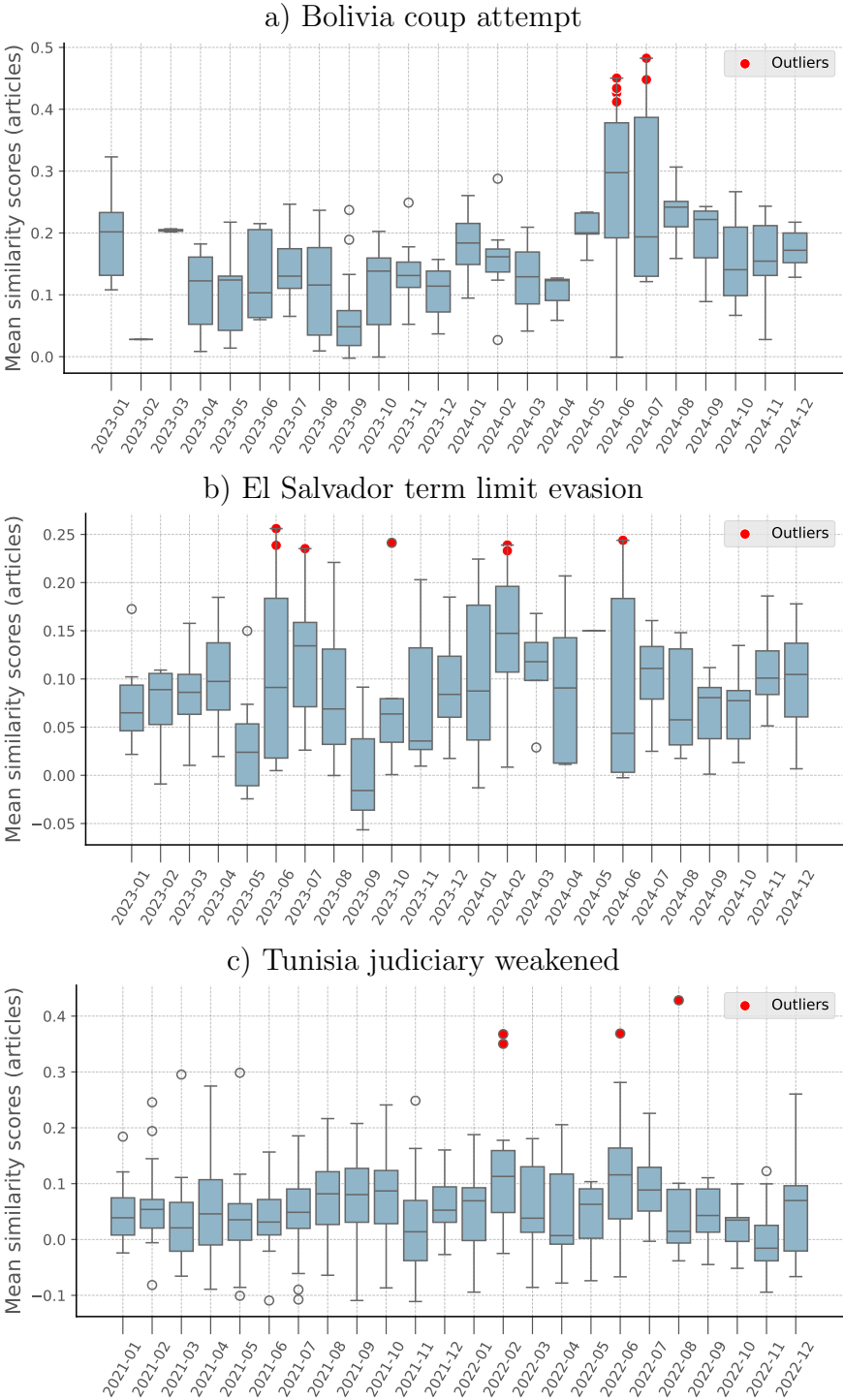
Figure 5 illustrates the distribution of similarity scores around known events. The box plots shows clear spikes and outliers coinciding with major events reported in the text. Panel a) shows the June 2024 failed coup attempt in Bolivia, included in the Cline dataset, where news coverage continues in the following month. Panel b) corresponds to El Salvador’s President Bukele beginning his second term in June 2024; additional outliers mark earlier milestones such as his nomination, registration, and election victory. Panel c) tracks Tunisia’s political crisis: the February 2022 dissolution of the High Judicial Council and the April 2022 seizure of the electoral commission. These cases illustrate how the method flags elevated similarity scores during and immediately after institutional disruptions, capturing both the main events and accompanying developments reflected in news reporting.

Figure 4: Event study plots of similarity scores before and after events



Notes: The coefficients and their 95% confidence intervals are obtained by regressing the mean similarity score of each event on dummy variables representing the period of the event and the 6 months before and after. The x-axis indicates the months before and after the event. The vertical red dashed line indicates when the event took place.

Figure 5: Case studies of distributions of similarity scores around events



Notes: Box plots show the median (central line), the interquartile range (box spanning the 25th–75th percentiles), and the whiskers, which extend to $1.5 \times$ the interquartile range. Points beyond the whiskers represent outliers. Panel a) show the distribution of headlines around the Bolivian coup attempt in June 2024, panel b) around the attempt to term limit evasion in El Salvador leading to Bukele rerunning in June 2024, and panel c) around attacks on the judiciary in February and April 2022.

4 Event detection: a nowcasting approach

When analyzing episodes of autocratization, we face two main challenges: the absence of an official and comprehensive database, and the need for constant updates at a frequency higher than annual. Our goal, therefore, is to leverage daily news information to construct and update monthly outcomes of democratic backsliding.

For the nowcasting exercise we use a supervised learning approach at the country-month level.⁸ In addition to including the average similarity scores for each month and country, the model incorporates the top-1 through top-5 scores, the standard deviation, and the maximum and minimum scores within the month. This design enables the classifier to be robust to variations in the volume of news across countries and months. We experimented with algorithms capable of handling missing values, and found that boosting tree models, namely CatBoost, LightGBM and XGBoost (shown in Table 3) delivered the best results.⁹ For the final XGBoost model we use 500 trees with a depth of 4 as weak learners. We set the learning rate to 0.01 and use early stopping to prevent overfitting.

To evaluate its out-of-sample performance, the nowcast is trained with data from 1989 up to and including 2014 and then tested on data up to December 2024. Table 3 presents the XGBoost nowcast performance metrics for detecting coups, term limit evasions and weakening of the judiciary.

A comparison with traditional dictionary-based approaches demonstrates clear improvements. Across all event types, the nowcast model dominates both simple keyword dictionaries and raw similarity-score thresholds, achieving higher true-positive rates at every

⁸To detect atypical events, we explored two methods. The first relied on a clustering algorithm, which identified clusters based on two dimensions: a spatial dimension, determined by high semantic similarity, and a temporal dimension, defined by the proximity of two or more high-score articles within a short time frame. While this method performed well with balanced data—where the number of monthly articles per country was relatively homogeneous—it faced significant limitations in our case. For instance, some countries average 500 monthly articles, while others have as few as 5. This imbalance required country-specific parameter adjustments, a costly process that could change as more data becomes available.

⁹CatBoost, LightGBM, and XGBoost are all boosting tree algorithms that construct predictions by combining many simple decision trees, each correcting the errors of the previous ones. Rather than imposing a linear or parametric structure, these models learn flexible nonlinear relationships and interactions directly from the data, which is particularly useful in settings with complex signals and heterogeneous effects. XGBoost is especially well suited to our application because it can handle large, high-dimensional prediction problems efficiently while flexibly capturing nonlinearities and interactions among covariates. It accommodates missing values in a systematic way through default split directions and includes strong regularization tools—such as shrinkage, tree-depth constraints, and subsampling—that help limit overfitting in settings with rare events and imbalanced outcomes. These properties make XGBoost well adapted to high-frequency, country-month panel data where predictors differ in scale, coverage, and predictive content across countries and over time.

false-positive level. For coups, the ROC curve shows an AUC of 0.90 for the nowcast, compared to 0.87 for the dictionary method and 0.82 for mean similarity scores. For term-limit evasions, the gap is more pronounced: the nowcast achieves an AUC of 0.92, while dictionary methods reach only 0.62. Judiciary-weakened events exhibit a similar pattern, with the nowcast clearly outperforming the dictionary approach (AUC = 0.80 vs. 0.53). Precision–recall curves tell the same story: the nowcast consistently achieves higher precision at any given recall level across all categories. These comparisons highlight that embedding-based text features coupled with supervised learning substantially outperform dictionary methods, particularly for institutional disruptions that are described in diverse ways across countries and often without standardized terminology.

Table 3: Out-of-sample nowcast performance metrics for autocratic events

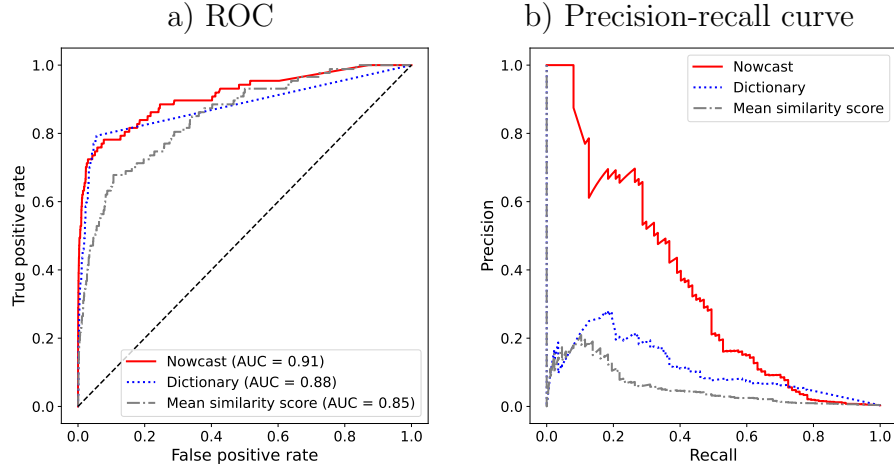
Target	ROC-AUC	PR-AUC	Precision	Recall	F1-Score
Coup	0.908	0.347	0.484	0.365	0.416
Term limit evasion	0.923	0.293	0.611	0.22	0.324
Judiciary weakened	0.784	0.121	0.369	0.146	0.21

Notes: Trained on 1989m1 to 2014m12 and tested on 2015m1 to 2025m10.

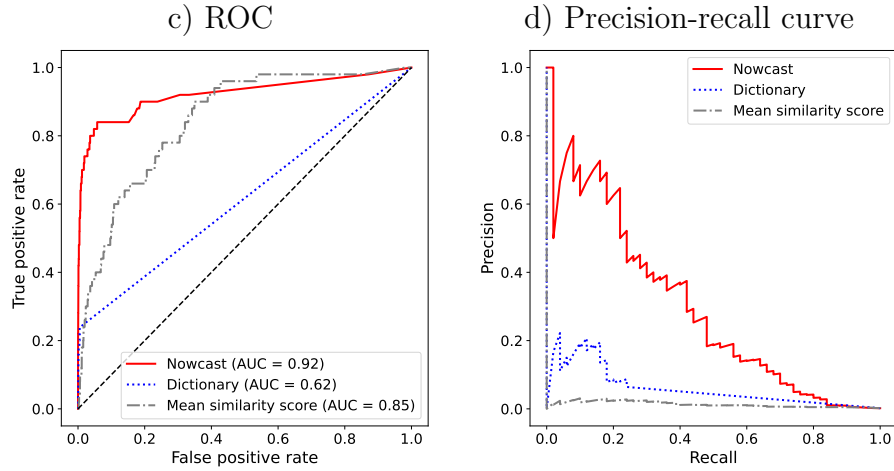
A comparison with traditional dictionary-based approaches in Figure 6 demonstrates clear improvements. Across all event types, the nowcast model dominates both simple keyword dictionaries and raw similarity-score thresholds, achieving higher true-positive rates at every false-positive level. For coups, the ROC curve shows an AUC of 0.90 for the nowcast, compared to 0.87 for the dictionary method and 0.82 for mean similarity scores. For term-limit evasions, the gap is more pronounced: the nowcast achieves an AUC of 0.92, while dictionary methods reach only 0.62. Judiciary-weakened events exhibit a similar pattern, with the nowcast clearly outperforming the dictionary approach (AUC = 0.80 vs. 0.53). Precision–recall curves tell the same story: the nowcast consistently achieves higher precision at any given recall level across all categories. These comparisons highlight that embedding-based text features coupled with supervised learning substantially outperform dictionary methods, particularly for institutional disruptions that are described in diverse ways across countries and often without standardized terminology.

Figure 6: Nowcast evaluation

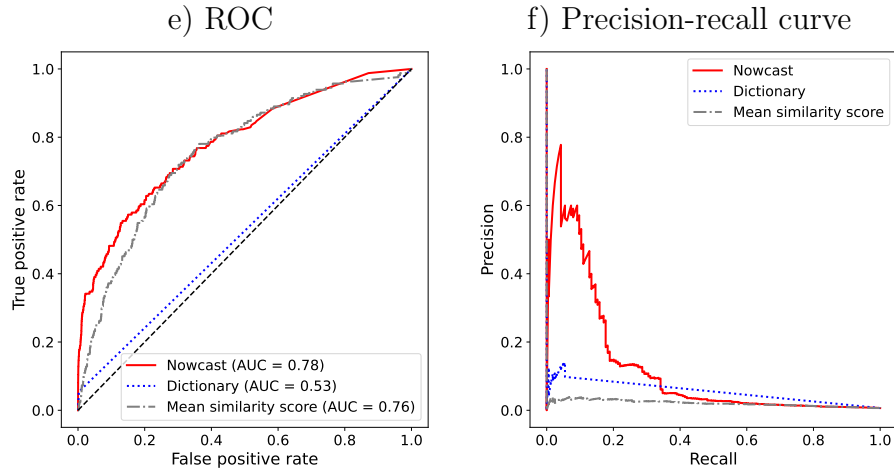
Panel A: Coups



Panel B: Term limit evasion



Panel C: Judiciary weakened

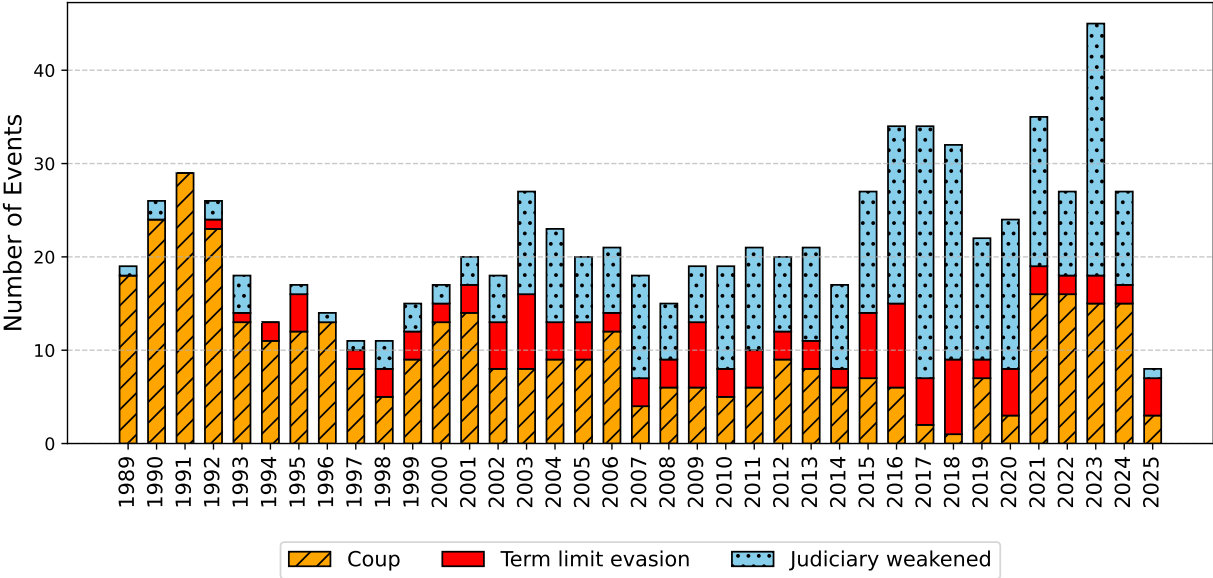


Notes: The solid line represents the performance of the nowcast obtained using XGBoost with features based on similarity scores as described in Section 4. The dictionary-based ranking is obtained using the terms specified in Appendix A.2. The dotted line represents the score obtained by computing the mean similarity score of the cosine similarity of the daily headlines to the event example headlines. The nowcast was trained with data from 1989m01 to 2014m12 and the results above are for the true out of sample test set up to 2025m10.

5 Predicting future events

After having generated the database of events we can now predict whether events will happen in the future. As shown in Figure 7, there were more coups in the 1990s and more events of judiciary weakening in the 2010s. The total number of events ranges from 11 in 1997 and 1998 to 35 in 2021.¹⁰

Figure 7: Overview of events by year



Notes: The x-axis shows the year and the y-axis the cumulative number of events in each year. The orange dashed bars represent coups, the red bars term limit evasions, and the blue dotted bars weakened judiciary.

For the prediction task, we rely on a Random Forest (Breiman, 2001) and implement an expanding window forecast methodology. The Random Forest operates by constructing multiple decision trees during training, each based on a random subset of the data and features. These individual trees are then combined to make predictions. Random Forest offers several advantages, such as handling high-dimensional data and nonlinear relationships, reducing overfitting, and providing feature importance rankings.

The following sets of features are used for each country-month:

- **Historical:** months since last disruption event, number of disruption events in the last 6/12/60/120 months, months since any fatalities recorded by the Upsala Conflict

¹⁰Appendix Figure A1 shows that a handful of the events overlap.

Data Program (UCDP), months since last election, number of elections within the last 12 months.

- **Topic-based:** monthly news volume (number of tokens), cumulative news volume (stock of tokens) with a discount factor for tokens further in the past, topic shares derived from LDA, topic stocks (also discounting the past).
- **Semantic similarity:** minimum, maximum and mean similarity scores of articles in that month.
- **Slow moving:** population.
- **Additional:** months until next election.

In our expanding forecast, the forecast horizon remains fixed (in our case 12 months into the future), but the forecast is updated at monthly intervals. As each period passes, we add the most recent actual data and update the forecast for the next prediction horizon.

To solidify the expanding window forecast methodology, we provide an example of the algorithm employed where the full data sample ranges from Jan 1989 to Aug 2023. This can obviously be modified by updating Aug 2023 to the latest available date for which data is available:

Algorithm 1: Expanding window forecast: Pseudo out of sample forecasting

Require: Full data sample $D = \{d_{1989m1}, d_{1989m2}, \dots, d_{2023m8}\}$

Require: Window size W

Require: Forecasting model F

Ensure: Forecasts $\hat{Y} = \{\hat{y}_{2010m1 < t \leq 2010m1+W}, \hat{y}_{2010m2 < t \leq 2010m2+W}, \dots, \hat{y}_{2023m8 < t \leq 2023m8+W}\}$

- 1: Train model F on data $\{d_{1989m1}, d_{1989m2}, \dots, d_{2009m12}\}$
 - 2: Optimize and fix hyperparameters of F using cross-validation on data $\{d_{2010m1}, d_{2010m2}, \dots, d_{2014m12}\}$
 - 3: **for** T from 2010m1 to 2023m8 **do**
 - 4: $D_{\text{train}} \leftarrow$ Data for all $1989m1 \leq t \leq T - W$
 - 5: Retrain model F on D_{train}
 - 6: $\hat{y}_t \leftarrow$ Aggregate forecast of F for $T < t \leq T + W$
 - 7: Append \hat{y}_t to \hat{Y}
 - 8: **end for**
- return** \hat{Y}
-

Using an expanding-window design, we generate a full history of pseudo-out-of-sample forecasts that replicate the information set available to a real-time decision maker. At each month T , the model is trained only on data observed up to T and produces a single prediction for the probability that an event occurs at any point in the next twelve months,

$T < t \leq T + W$ with $W = 12$, rather than forecasting each future month separately. For example, rewinding to 2015m1, a policymaker would only have access to data through 2014m12 when forming a forecast for 2015m1–2015m12, and our procedure mirrors this timing. This provides a realistic benchmark for forecasting performance without contaminating the test set with future information. However, results should be interpreted as an upper bound, since the overall modeling strategy reflects researcher choices that may inflate performance relative to real-time use. ¹¹

We train a model to learn a functional form using all data from 1989m1 to 2010m1 as follows:

$$y_{i,T < t \leq T+W} = F_T(\mathbf{X}_{i,T}),$$

With the resulting model, we then produce out of sample predictions on a rolling basis from 2010m1 onwards:

$$\hat{y}_{i,T < t \leq T+W} = F_T(\mathbf{X}_{i,T}),$$

Hyperparameters are chosen by maximizing the area under the curve (AUC) of the receiver operating characteristics (ROCs) curve via pseudo-out-of-sample rolling forecasting on the sample 2010 to 2015. The final hyperparameters of the Random Forest are characterized by a tree depth of 100 and a restricted tree complexity with a maximum depth of 6, limiting the scope for overfitting. At each split, only 10% of available predictors are randomly considered, corresponding to roughly 5–6 features, which increases model diversity and reduces reliance on any single covariate. Finally, we impose a large minimum leaf size of 150 so that terminal nodes contain sufficient observations to deliver stable probability estimates, an important consideration given that disruptive events are rare.

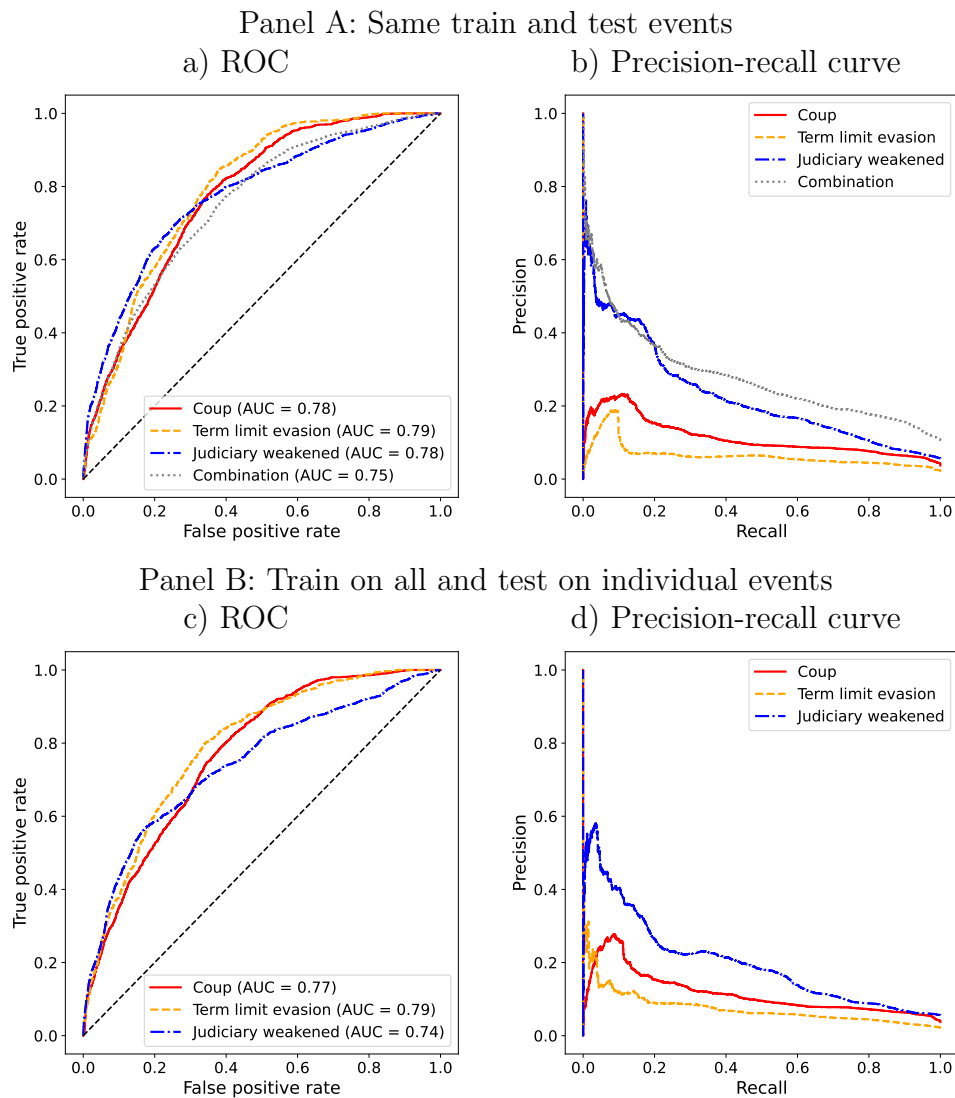
5.1 Results

Figure 8 summarizes the pseudo-out-of-sample forecasting performance of our models using an expanding window from 2010m1 to 2024m10. Across all event categories, i.e. coups, term limit evasion, and judicial weakening, as well as their combination, the ROC curves display similar shapes. The corresponding AUC values lie in a narrow range between roughly 0.74 and 0.79, indicating a stable ability of the models to rank months with higher risk above months with lower risk across different institutional outcomes. This similarity suggests that the underlying predictors capture a common component of political instability that is relevant

¹¹We describe this as an upper bound because, even though the test set is not used for estimation, the overall modelling strategy—including feature choices and algorithm selection—has been chosen to achieve the best result. This type of “researcher overfitting” tends to inflate out-of-sample performance relative to what a policymaker could achieve in real time without the benefit of hindsight or trial-and-error on the same dataset.

across event types, rather than being narrowly tailored to a specific outcome.

Figure 8: Forecast evaluation



Notes: Rolling forecast trained from 1989m1 onward and evaluated from 2010m1 to 2024m10.

However, as Panel A also makes clear, ROC curves alone can mask economically meaningful differences when events are rare. The precision–recall curves therefore provide a different perspective on operational forecasting performance. Because precision explicitly accounts for class imbalance, it declines sharply for the rarer events, most notably term limit evasion, despite comparatively strong ROC-AUC values. This divergence highlights that good ranking performance does not necessarily translate into a high fraction of correct alarms once a decision threshold is imposed.

The combined-events specification, predicting the union of all three disruptive events, illustrates this tradeoff particularly clearly. When aggregating coups, term limit evasion, and judicial weakening into a single outcome, the model achieves a substantially higher precision at moderate levels of recall. For example, at a recall of about 50%, precision rises to roughly one third, meaning that one out of three alarms corresponds to an actual event. From a policy or early-warning perspective, this represents a meaningful improvement relative to forecasting individual event types in isolation, where precision is often well below 20% at similar recall levels. The improvement reflects the fact that aggregation mitigates extreme class imbalance and leverages shared predictive signals across institutional breakdowns.

Table 4 reinforces these patterns in a summary-metric format. When models are trained and tested on the same event type, ROC-AUC values are uniformly high, but average precision varies widely. Judicial weakening stands out with comparatively high average precision despite a lower ROC-AUC for “hard” onsets, defined as five years without the corresponding event, suggesting that while the model may struggle to sharply separate the most difficult cases, it nonetheless produces relatively reliable positive signals when it does issue warnings. In contrast, term limit evasion exhibits strong a ROC-AUC but very low average precision, underscoring the difficulty of translating ranking performance into actionable forecasts for extremely rare events.

The bottom panel of the table further shows that training on the pooled set of events does not materially degrade performance when forecasting individual outcomes. ROC-AUC values remain close to those obtained from event-specific training, while average precision is often comparable or slightly improved. This finding suggests that pooling events is not merely a statistical convenience but can be justified substantively: different manifestations of democratic erosion appear to share a common latent risk structure that can be exploited for forecasting through transfer learning.

In Appendix Figure A3 we give an overview of which predictors contribute most to the forecast of future events using a beeswarm plot of SHAP values. The best predictors of future disruptive events are past disruptive events. However, embedding similarities between emerging headlines and the prototype headlines clearly add to the overall predictive power of the model. Finally, as in conflict prediction, we find that LDA topics of the overall news coverage in a given country also contributes to the prediction of events at the margin.

Table 4: Pseudo-out-of-sample forecast performance metrics

Train data	Test data	ROC-AUC		Avg
		Overall	Hard	Precision
Coup	Coup	0.775	0.773	0.113
TLE	TLE	0.794	0.784	0.064
JW	JW	0.776	0.683	0.226
Coup + TL + JW	Coup + TL + JW	0.752	0.699	0.279
Coup + TL + JW	Coup	0.769	0.756	0.113
Coup + TL + JW	TLE	0.791	0.769	0.078
Coup + TL + JW	JW	0.742	0.614	0.197

Notes: Rolling forecast trained from 1989-01 onward and evaluated from 2010m1 to 2024m10 (the last month for which we can evaluate the next 12 months).

5.2 Future outlook

Table 5 illustrates the forward-looking risk assessment generated by our forecasting system with data available until October 2025, highlighting countries with the highest predicted likelihood of institutional disruptions over the next twelve months. The rankings reveal substantial heterogeneity across both countries and types of events. For example, several countries that appear near the top of the coup-risk column do not rank highly in terms of term-limit evasion or judicial weakening, indicating distinct underlying dynamics. Conversely, some countries with elevated risks of judicial weakening or term-limit manipulation show little immediate risk of a coup, consistent with more gradual forms of executive overreach. The combined risk ranking aggregates these dimensions and highlights countries where any form of institutional pressure may be present.

Since these forecasts were produced, Benin experienced a coup attempt in December 2025. While Benin did not rank as high in the predicted likelihood of experiencing a coup, it did rank 11th in the likelihood of experiencing at least one of the three disruptive events, with a predicted probability of 22.8%. This illustrates that forecasting rare events is inherently a search for a needle in a haystack, but the model nonetheless succeeds in flagging countries facing elevated institutional risk, helping to narrow attention toward a manageable set of plausible candidates.

6 Impact study

In Figure 9, we plot the dynamic treatment effects of coups d'état on log GDP per capita using two alternative empirical approaches. The horizontal axis shows years relative to the

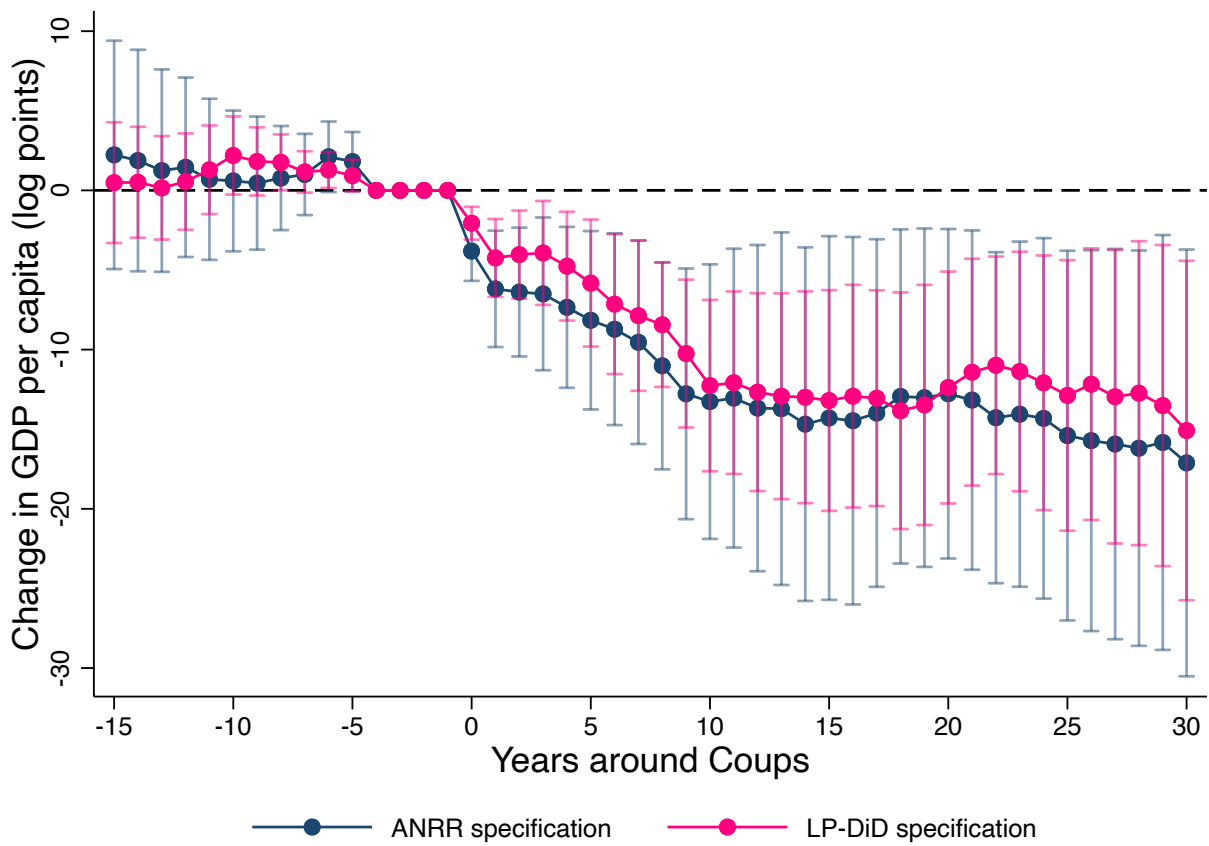
Table 5: Countries with highest event risk

Rank	Coup		Term limit evasion		Judiciary weakening		Combined	
	Country	Risk	Country	Risk	Country	Risk	Country	Risk
1	GNB	0.189	SSD	0.108	BRA	0.243	SLV	0.359
2	MDG	0.168	DJI	0.083	MEX	0.223	GTM	0.298
3	BFA	0.149	TCD	0.077	PER	0.214	BRA	0.279
4	COD	0.146	MLI	0.074	ARG	0.205	GIN	0.273
5	HTI	0.137	SLV	0.074	GTM	0.172	GNB	0.267
6	SDN	0.126	COD	0.073	SLV	0.147	MEX	0.26
7	BLR	0.121	SOM	0.064	BOL	0.144	MDG	0.255
8	SYR	0.117	CAF	0.049	ECU	0.136	COD	0.24
9	UKR	0.115	CIV	0.048	GIN	0.127	PER	0.237
10	NER	0.112	ECU	0.044	BEN	0.113	SSD	0.236

Notes: Forecast of events within the next 12 months, generated with data up to and including 2025m10.

coups, with year 0 denoting the coup year, while the vertical axis reports changes in GDP per capita in log points. Both the event-study / difference-in-differences specification following [Acemoglu et al. \(2019\)](#) and the local projections difference-in-differences estimates following [Dube et al. \(2025\)](#) display no systematic pre-trends in the years leading up to coups, suggesting the absence of differential growth dynamics prior to the event. Following a coup, GDP per capita declines sharply and persistently. The estimated effects reach around 10–15 log points within the first decade after the coup and continue to deteriorate over longer horizons. While confidence intervals widen at longer horizons, both specifications point to substantial and long-lasting economic losses associated with coups, with broadly similar dynamics across methods.

Figure 9: Treatment effects of Coups d'État on the log of GDP per capita



Notes: The results are obtained following the specification from Section IV of [Acemoglu et al. \(2019\)](#) (ANRR) and the local projections difference-in-differences approach proposed by [Dube et al. \(2025\)](#) (LP-DiD). Solid points represent the estimated average effects, with bars indicating 95% confidence intervals.

7 Conclusion

This paper develops and validates a scalable system to detect and predict institutional disruptions, including coups d'état, term-limit evasion, and the weakening of the judiciary, by combining established event datasets with information extracted from a large global news corpus. Our contribution is threefold. First, we construct a regularly updatable database of disruption events by nowcasting incidents from embedding-based semantic similarity signals and subjecting the resulting candidates to a human-in-the-loop verification step before integrating them with canonical sources. Second, we show that these enriched event targets support meaningful out-of-sample forecasting performance, including for hard onsets that occur after long periods of institutional stability. Third, we provide an operational monitoring tool that produces monthly, low-marginal-cost risk assessments over a 12-month horizon for more than 170 countries, enabling near-real-time tracking of institutional stress.

Methodologically, the paper illustrates how modern sentence embeddings and semantic similarity measures can be used to expand sparse event labels when hand-coding is costly or slow, and to generate informative, high-frequency predictors for forecasting. Aggregating article-level signals to the country-month level mitigates cross-country heterogeneity in news volume, while our expanding pseudo-out-of-sample evaluation design aligns model assessment with the information set available to real-world decision makers. The ability of a single forecasting system to learn from multiple forms of institutional disruption further suggests that autocratization is best viewed as a set of related processes with shared informational signatures, rather than as isolated event types.

At the same time, the approach has important limitations that motivate further research. News data are subject to unequal coverage, censorship, and reporting incentives, and measurement error remains a concern in both text-based signals and legacy event labels. The rarity of institutional breakdowns also raises challenges for calibration and threshold choice in applied early-warning settings. Addressing these issues, through improved bias correction, alternative validation strategies, and more explicit uncertainty quantification, represents a natural next step for this research agenda.

Despite these caveats, the framework provides a practical way to identify where and when institutional risks intensify, at a frequency and transparency that traditional datasets alone cannot deliver. By producing traceable and reproducible disruption signals, it can help researchers, civil society, and policymakers allocate scarce attention, plan preventive engagement, and evaluate the consequences of institutional stress. Moreover, the resulting event database, with clearly defined event types, and verifiable text evidence, supports both

qualitative case study work and large-scale empirical analysis. As an illustration of its economic relevance, we document that coups are followed by large and persistent declines in economic growth. Future work can build on this infrastructure to study mechanisms and strengthen causal interpretation, for example by combining these high-frequency measures with instrumental-variable designs or structural modeling approaches.

References

- Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J. A. (2019). Democracy does cause growth. *Journal of Political Economy*, 127(1):47–100.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- Boese-Schlosser, V. and Eberhardt, M. (2025). Democracy in decline: The economic implications of democratic collapse. CEPR Discussion Paper 20146, CEPR Press, Paris & London. DP20146.
- Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.
- Democratic Erosion Consortium (2025). *Democratic Erosion Event Dataset Codebook v7*.
- Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.
- Dube, A., Girardi, D., Jorda, O., and Taylor, A. M. (2025). A local projections approach to difference-in-differences. *Journal of Applied Econometrics*, 40(7):741–758.
- Funke, M., Schularick, M., and Trebesch, C. (2020). Populist leaders and the economy. CEPR Discussion Paper 15405, CEPR. DP15405.
- Ginsburg, T., Melton, J., and Elkins, Z. (2011). On the evasion of executive term limits. *William & Mary Law Review*, 52(6):1807–1872.
- Gottlieb, J., Balvanera, R., Baron, H., Blair, R. A., Paler, L., and Weaver, J. A. (2025). Democratic erosion event dataset v7.
- Helmke, G. (2017). *Institutions on the Edge: Inter-Branch Crises in Latin America*. Cambridge University Press.
- Hoffman, M., Bach, F., and Blei, D. (2010). Online learning for latent dirichlet allocation. *advances in neural information processing systems*, 23.
- Juhász, R., Lane, N. J., Oehlsen, E., and Perez, V. C. (2025). Measuring industrial policy: A text-based approach. Technical report, National Bureau of Economic Research.
- Mueller, H. and Rauh, C. (2022). The hard problem of prediction for conflict prevention. *Journal of the European Economic Association*, 20(6):2440–2467.
- Mueller, H., Rauh, C., and Seimon, B. (2024). Introducing a global dataset on conflict forecasts and news topics. *Data & Policy*, 6:e17.
- Peyton, B., Bajjalieh, J., Shalmon, D., Martin, M., Bonaguro, J., and Soto, E. (2023). Cline center coup d’État project dataset.
- Powell, J. M. and Thyne, C. L. (2011). Global instances of coups from 1950 to 2010: A new dataset. *Journal of Peace Research*, 48(2):249–259.
- Razin, A. (2024). The transition to illiberal democracy: Economic drivers and consequences. *CEPR e-book*.
- Řehřek, R. and Sojka, P. (2010). Software framework for topic modelling with large corpora.
- Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American economic review*, 94(4):1055–1084.

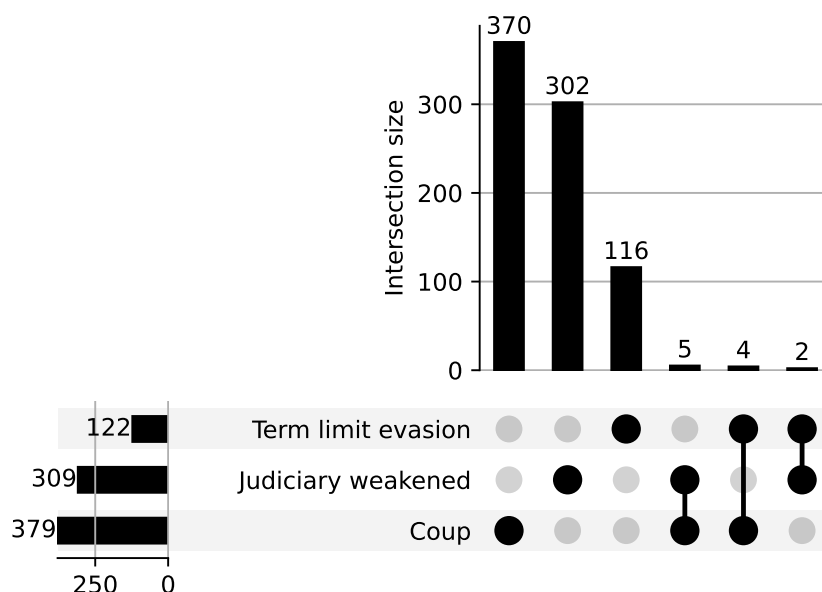
- Versteeg, M., Horley, T., Meng, A., Guim, M., and Guirguis, M. (2020). The law and politics of presidential term limit evasion. *Columbia Law Review*, 120(1):173–248.
- Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., and Zhou, M. (2020). Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. Wenhui Wang.

A Appendix

A.1 Event overview

Coups are the most frequent event type, followed by cases of term limit evasion and weakening of the judiciary. Only 11 country-months see a co-occurrence of two event types. For example, Sri Lanka going through a constitutional crisis (under President Maithripala Sirisena) in November 2018 is included both by the Cline data set as a coup attempt and by the DEED data as a case of ‘delegitimizing or weakening judiciary’. Meanwhile Honduras in June 2009 saw President Zelaya attempt to extend term limits, which is one of the reasons he was ousted in a coup later that same month.

Figure A1: Illustration of overlapping events



Notes: This upset graph shows how many events there are of each category and counts instances where more than one event takes place in single country-month unit.

A.2 Dictionaries for baseline model

These dictionary terms were used to generate a baseline of how to rank country-months by their likelihood of experiencing a given form of institutional disruption. At the article level, we use regular expressions to capture whether at least one of the terms for each event type appears in the title. To aggregate this at the monthly level, we divide the number of articles containing relevant terms by the total number of articles for each country-month.

Regular expressions for coup:

- \bcoup\b
- \bseize (power|control)\b
- \bousted\b
- \btoppled\b
- \bmartial law\b
- \bmilitary rule\b
- \bmilitary stages\b

Regular expressions for term limit evasion:

- \b(abolish|remove|scrap|end|overturn)(s|es|ed)? (presidential)?term limit(s)?\b
- \bextend(ing|ed|s)? (presidential)?term\b
- \bpresident for life\b
- \blife presidency\b
- \bseek(s|ing)? (a(nother)?|second|third) term\b
- \b(third|fourth|fifth|3rd|4th|5th) term\b
- \bage limit lifted\b
- \bconstitutional coup\b
- \blawmakers (vote|pass|approve) (to|for)? .*term\b
- \bcourt (clears|validates|says) .* (run|term)\b
- \breferendum (passes|approved)\b
- \bclashes over term limits\b
- \bprotests against longer rule\b
- \bdemonstrators protest\b
- \bunconstitutional\b
- \b(remains? in|clings? to) power\b

Regular expressions for judiciary weakened:

- \b(crackdown|attack) on (judiciary|courts)\b
- \bcourt (overhaul|reconfiguration|realignment)\b
- \bjudicial purge\b
- \bjudges (dismissed|replaced|forced to retire)\b
- \bloyal(ists| judges?)? appointed\b
- \bchief justice (removed|appointed|ousted|resigns)\b
- \bcontroversial (nominee|appointments?)\b

- \bappoint(s|ed)? allies\b
- \bconstitutional (court powers|amendments)\b
- \blimits? on judicial review\b
- \b(curtail|limit) court oversight\b
- \breform judiciary\b
- \b(restrict|threats? to) judicial independence\b
- \bjudicial independence (at risk|questioned)\b
- \balter court powers\b
- \bexecutive overreach\b
- \bpower grab\b
- \bbias(ed)? judges\b
- \baccused of (influence|interference)\b
- \blegal community warns\b

A.3 Support set (example headlines)

Coup

index	Example headlines: coup
0	Breaking: Coup attempt reported in the capital
1	Authorities confirm coup d'etat underway, leaders claim control
2	Military stages coup d'etat, seizing key government buildings
3	Coup attempt fails as loyalists regain control of key assets
4	Coup leaders announce new regime, suspending the constitution
5	Tensions rise as coup forces take control of national broadcasting
6	Coup attempt unfolds as military clashes with government forces
7	Reports of a coup attempt emerge as leaders call for calm
8	Coup leaders declare victory, impose curfew across the city
9	Coup d'etat results in the suspension of parliament sessions
10	Government denounces coup attempt, vows to defend the constitution
11	Coup forces seize control of national landmarks, curfew imposed
12	Reports of coup leaders holding meetings with foreign emissaries
13	Coup d'etat triggers unrest, clashes erupt in the central district
14	Coup attempt crushed by government loyalists, leaders captured
15	Coup d'etat sees new leaders emerge, old regime ousted
16	Tanks seen in the capital as coup leaders make power move
17	Coup leaders establish temporary military rule, calls for elections
18	Coup attempt sparks protests, citizens demand democratic restoration
19	Military coup leaders announce takeover of vital industries
20	Coup d'etat declared successful as military takes control

index	Example headlines: coup (cont.)
21	Coup attempt thwarted, military returns power to elected leaders
22	Coup forces declare martial law, restrict civil liberties
23	Coup d'état raises concerns over stability in the region
24	Coup attempt leads to power vacuum, opposition gains momentum
25	Leaders of the coup declare state of emergency, control media outlets
26	Coup attempt organizers apprehended, government regains control
27	Coup d'état creates uncertainty, protests break out across the capital
28	Coup forces broadcast demands for political reforms over state TV
29	Coup attempt sees rising tensions between military and government
30	Military coup d'état leaves power in the hands of commanders
31	Coup leaders shut down internet, impose nationwide curfews
32	Coup d'état disrupts transportation and communication systems
33	Citizens in shock as coup leaders take over national media
34	Coup attempt falters, military withdraws to bases
35	Coup forces announce transition government, promise swift reforms
36	Coup attempt marks a new chapter of instability in the region
37	Coup leaders consolidate power, dismiss former government officials
38	Coup d'état triggers international concerns over human rights abuses
39	Coup attempt results in temporary suspension of civil rights
40	Military coup leaders plan to hold talks with opposition forces

Term limit evasion

index	Example headlines: term limit evasion
0	abolishes presidential term limits
1	Legislature Ends Presidential Term Limits
2	approves constitution changes, reappoints premier
3	parliament OKS new constitution with term limits
4	new constitution likely to have "negative effect" on politics - pundit
5	parliament rejects bill to reduce president's term
6	Après moi, moi; president v term limits
7	court validates president's third term bid
8	constitutional court okays president's re-election bid
9	secretary of state urges leader to respect constitution
10	President Says Third Term Would Be His Last
11	Leader Can Extend His Term. His African Peers Take Notes.
12	New constitution gives president absolute powers'
13	Leader Aims to Remain in Power After 3 Terms
14	Voters Back Constitution Extending Presidential Term
15	Opposition says controversial amendment "constitutional coup"
16	Constitutional referendum to remove presidential term limits divides Republic
17	Opposition MPs walk out as parliament adopts constitutional amendment
18	Leader Backed For Third Term
19	votes for new constitution giving president more power

index	Example headlines: term limit evasion (cont.)
20	referendum approaches as opposition calls for boycott
21	Referendum: Coups, I did it again
22	Military deploys across capital as President stays on
23	Clinging to Power
24	President Clings to Power
25	lawmakers vote to end presidential term limit
26	Speaker defends army's role in amended constitution
27	Speaker says amendments 'not to make power permanent'
28	MPs approves extending presidential term to six years
29	schedules vote on constitutional changes aimed at keeping President in power until 2030 for April 20-22
30	opposition urges "No" vote to extend president's term
31	Opposition official says constitution review bill "unacceptable"
32	Referendum approves constitutional amendments
33	opposition rejects new constitution
34	high court voids ban on presidential re-election
35	Term Limits Overturned
36	electoral commission says that President has overwhelmingly won a third term
37	President Wins Unconstitutional 3rd Term in Disputed Vote
38	gives 30-day deadline to restore constitutional order
39	opposition demonstrates to protest coup
40	Court orders electoral body to opposition presidential candidate
41	President Announces He Will Seek Second Term
42	President announces re-election bid opponents call illegal
43	Leader to Seek a 3rd Term, Capping a Long Legal Battle
44	Demonstrators protest Peru president's re-election bid
45	to hold referendum to remove presidential term limit
46	6 injured in clashes over term limits
47	4 killed in protests against longer rule
48	writer views constitution changes, expansion
49	Overwhelming majority vote in favour of constitutional changes
50	main law briefly published as President's constitution
51	Lawmakers Clear Referendum on Letting President Seek a Third Term
52	Opposition party terms leader's third term bid "danger fro democracy"
53	wants to set up a republican monarchy - former PM
54	express concern over on constitutional amendment
55	president cleared to run for 3rd term
56	Court Says Leader Can Run for Third Term
57	removes term limits for presidency
58	Parliament Ends Term Limits on the President
59	constitutional amendment new powers are unnecessary and dangerous
60	Opposition leader says constitutional amendment "a crime"
61	parliament approves presidential term limits
62	Referendum to extend president's term passes by more than 99 percent

index	Example headlines: term limit evasion (cont.)
63	votes to expand president's powers; critics cry fraud
64	main opposition party submits formal request seeking referendum to be annulled, citing irregularities
65	Third term bill sent to president for assent
66	leader supports removing presidential age limit
67	Lifts an Age Limit, Paving the Way for a President for Life
68	lawmakers pass contentious 'life presidency' bill
69	opposition protests ending term limits
70	Want More? Campaign on Term Limits Amid Rising Tension
71	backers rally to end term limits
72	wins vote to scrap term limits
73	Leader Is Nominated For 3rd Term Despite Dissent
74	MPs opposing presidential third term seek dissolution of parliament
75	Non-governmental organizations oppose constitutional amendments

Judiciary weakened

index	Example headlines: judiciary weakened
0	President Expands Control: Courts Declare Opposition Party Illegal
1	Emergency Powers Extended: Constitutional Court Backs Controversial Government Moves
2	Supreme Court Backs Martial Law Extension: Critics Fear Weakening of Checks and Balances
3	High Court Deemed 'Enemy of the State': Controversial Overhaul of Supreme Court
4	Court Overhaul Clears Path for Sweeping Economic Reforms
5	Judiciary Under Fire: Judges Resign as Government Replaces Supreme Court Members
6	President Reshapes Judiciary: Dismissals and Reappointments Raise Eyebrows Among Legal Experts
7	Judicial Overhaul Sparks Fears of Authoritarianism: Supreme Court Realigned
8	Judiciary Purge Continues: Government Installs Loyal Judges as Crackdown Deepens
9	President Expands Control Over Judiciary: New Law Alters Constitutional Court Powers
10	Judges Forced into Early Retirement: Critics Say Government is Targeting Independence
11	Constitutional Amendments Passed: Government Curtails Court's Oversight Role
12	Opposition Decries Power Grab: Parliament Weakens Judicial Review
13	New Chief Justice Appointed: Allies of President Fill Vacancies in Constitutional Court
14	President Purges Courts: Thousands of Judges and Prosecutors Dismissed

index	Example headlines: judiciary weakened (cont.)
15	Constitutional Court Reconfigured: Executive Gains Power Over Appointments
16	New Decree Restricts Judicial Independence: Critics Warn of Increasing Executive Overreach
17	Judicial Loyalty Oaths Introduced: President Tightens Grip Over Judiciary Amid Crackdown
18	Government Pushes Through Controversial Judicial Reforms: Countries Express Concern
19	Constitutional Tribunal Crisis: Ruling Party Replaces Judges, Limits Court's Role
20	Supreme Court Judges Forced to Retire: President Signs Controversial Law
21	Judicial Independence at Risk: Legal Experts Warn of Diminishing Checks and Balances
22	New Court Appointments Spark Protests: Ruling Party Fills Tribunal with Loyalists
23	Executive Criticism of Judiciary Intensifies: President Suggests Term Limits for Supreme Court Justices
24	Controversial Appointments Raise Concerns: Supreme Court Nominee Labeled 'Partisan Pick'
25	New Legal Doctrine Proposed: Limits on Judicial Review Powers Spark Constitutional Debate
26	Supreme Court Under Pressure: Critics Warn of Erosion of Judicial Independence
27	Supreme Court Judges Warned: Congress Debates Limiting Court's Powers
28	Constitutional Court Realigned: President Appoints Close Allies to Key Positions
29	Judges Ordered to Comply: Executive Demands Loyalty From Judiciary
30	Chief Justice's Impeachment Sparks Outcry: Legal Community Warns of Judicial Erosion
31	Protesters against judiciary overhaul urge president to 'respect democracy'
32	President Tightens Grip on Judiciary: Chief Justice Forced to Resign
33	Judicial Independence Questioned: Executive Calls for Investigation Into 'Biased' Judges
34	Senate Moves Quickly to Confirm Judges: Record Number of Appointments Under Current Administration
35	Government Faces Backlash Over Judicial Appointments: Chief Justice Demands Transparency
36	New Chief Justice Appointed Amid Controversy: Executive Accused of Undue Influence
37	Judges Protest Interference: Historic Public Statement by Four Senior Judges Sparks Debate
38	President Criticizes Supreme Court: Calls for Changes in Judicial Appointments Process

index	Example headlines: judiciary weakened (cont.)
39	Constitutional Court Judges Face Threats: Government Allies Demand Reforms
40	New Chief Justice Appointed Amid Tensions: President Ally Takes Helm of Supreme Court
41	Government Ousts Chief Justice: Supreme Court Under Siege

A.4 Embedding model

A critical component of our process is the embedding module, which can be developed using either traditional supervised learning or self-supervised learning techniques, such as BERT and its derivatives. Initially, we employed the BERT base model (Devlin et al., 2018), a widely used transformer model. However, BERT’s original design is not optimized for generating meaningful sentence embeddings, as it was primarily intended for specific downstream tasks rather than general-purpose sentence representation.

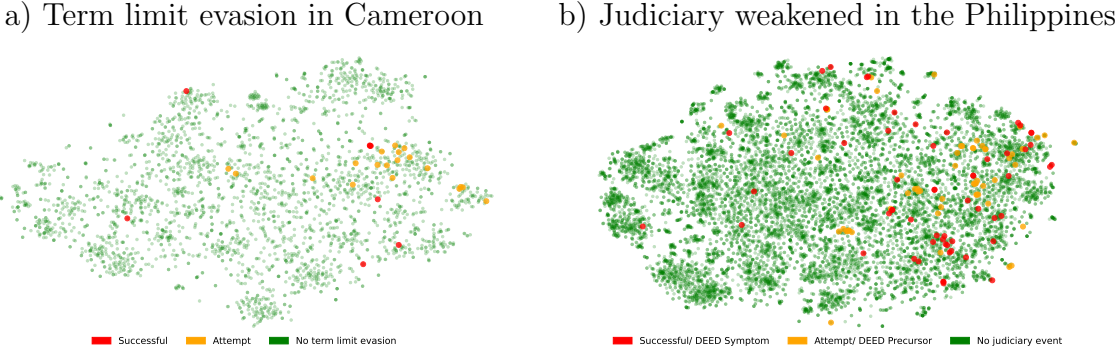
To address this limitation, Sentence Transformers introduce a pooling operation to BERT’s output, creating fixed-size sentence embeddings. This is further enhanced by fine-tuning the model using a triplet network, which improves the semantic alignment of the embeddings. The resulting model can process a list of sentences and generate semantically rich embeddings for each sentence.

In line with recent empirical research, we initially used the SBERT base model (Reimers and Gurevych, 2019). However, we found that this model is deprecated and tends to produce lower-quality sentence embeddings. Consequently, we are exploring alternative sentence embedding models to improve performance. Table A4 presents a comparison of these models, evaluating their performance on a simple example involving three queries and one prompt. The *all-MiniLM-L6-v2* (Wang et al., 2020) model demonstrates the best quality in terms of performance and time needed. Therefore we will switch to this transformer going forward.

Prompt	“Term-limit evasion”				
Queries	“Chavez wins vote to scrap term-limits in Venezuela”				
	“I have to express myself in better terms”				
	“President Says Third Term Would Be His Last”				
Model	Similarity scores			Time taken (s)	Ranking
BERT	0.514	0.480	0.504	1.694	4
SBERT	0.149	0.245	0.123	2.145	5
all-mpnet-base-v2	0.292	0.128	0.270	1.664	3
all-distilroberta-v1	0.232	0.075	0.304	1.470	2
all-MiniLM-L6-v2	0.396	0.063	0.281	1.372	1

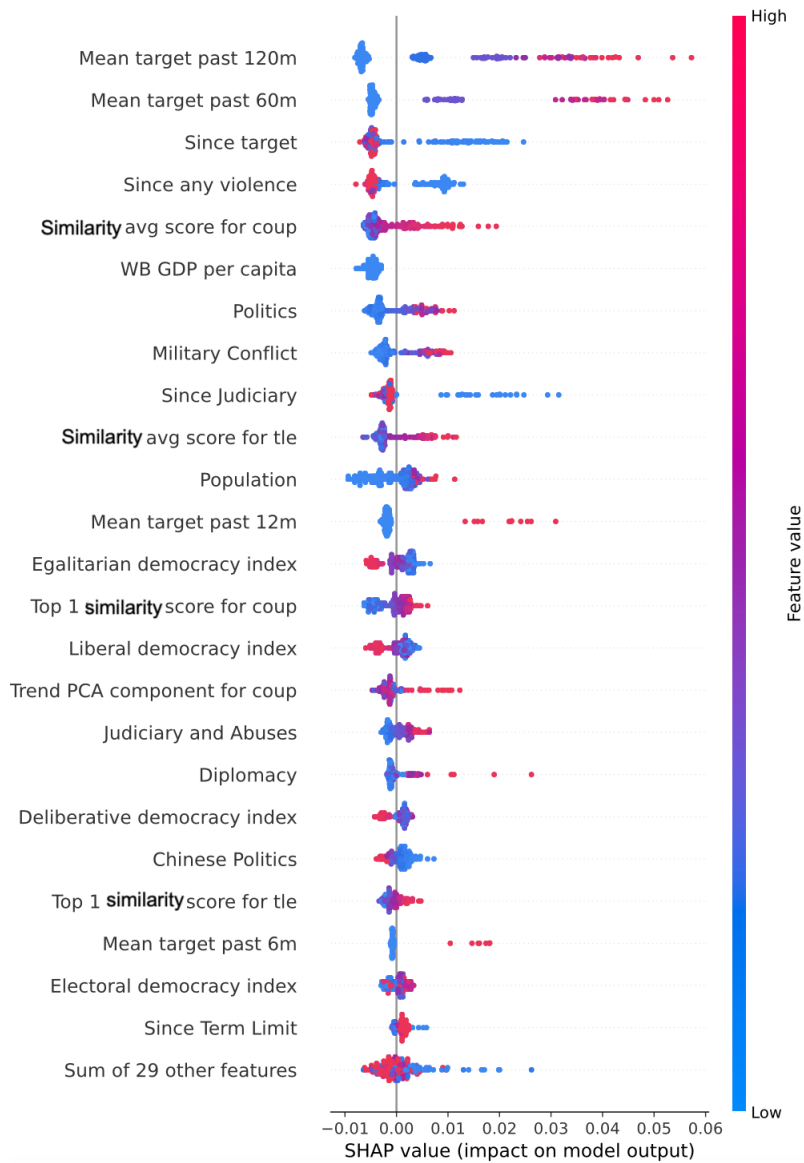
Table A4: Performance of Sentence Transformer Models. We loaded pretrained Sentence Transformer models and calculated the embedding similarities of each example query against the prompt. The pretrained models were obtained from the [Hugging Face organization](#).

Figure A2: Two-dimensional representation of article embeddings during term limit evasion and judiciary weakening



Notes: Notes: We use t-SNE to reduce the 384 dimensional embedding vectors of each headline pertaining to the Democratic Republic of the Congo into a two dimensional space. Red dots represent headlines around successful coups, orange around attempted coups, and black dots around coup conspiracies. The green dots represent the remaining headlines.

Figure A3: Beeswarm SHAP values when predicting union of events



Notes: The predicted outcome is any of the three events within the next twelve months in a country: judiciary weakening, term limit evasion, or coup. The prediction model is a random forest. The beeswarm plot visualizes feature importance using SHAP values, which decompose each model prediction into additive contributions of individual covariates. SHAP values are based on the Shapley value concept from cooperative game theory: each variable's contribution is defined as its average marginal effect on the prediction across all possible combinations of included variables. In the plot, each dot represents one observation. The horizontal position of a dot is the SHAP value for that feature (positive values increase the predicted event risk; negative values decrease it). Features are ordered by overall importance (mean absolute SHAP value). Dot colors indicate the underlying feature value (typically blue = low, red = high), allowing interpretation of whether higher or lower realizations of a predictor push predictions up or down. The vertical spread reflects heterogeneity across observations with similar importance.