MEASURING GEOPOLITICAL RISK WITH OPEN SOURCES: DATA FOUNDATIONS, CODING SCHEMES, AND CHALLENGES

https://doi.org/10.59982/18294359-25.2-mg-20

Henrik Sergoyan
Post-Graduate Student
EUA, Chair of Information Technologies and Applied Mathematics
henriksergoyan@gmail.com

Abstract

This article reviews how geopolitical risk can be quantified using open-source intelligence (OSINT) drawn from news media, event databases, social platforms, and complementary public data. We first define geopolitical risk as uncertainty arising from conflict, terrorism, and interstate tensions, and then motivate its economic relevance. We then examine four practical quantification routes: (i) news-based indices that transform media coverage into time-series signals of risk (e.g., measures in the tradition of the GPR index); (ii) event databases such as GDELT and ACLED that code protests, violence, and diplomatic actions into structured records that can be aggregated into frequency and intensity metrics; (iii) social media and search trend analytics that act as early-warning signals through sentiment and volume shifts; and (iv) composite, multi-source indices (e.g., trade-risk and market-attention gauges) that integrate sanctions, deployments, disruptions, and text signals into a single score. Brief case examples illustrate how these approaches capture historical surges and support operational decision-making in public and private sectors. We also synthesize key limitations—data overload and verification, media and regional bias, complex causality, and timeliness trade-offs—and emphasize the role of human expertise alongside automated analytics. The paper concludes with implications for researchers and practitioners seeking transparent, reproducible, and decision-relevant measures of geopolitical instability.

Keywords: Geopolitical Risk, Open-Source Intelligence (OSINT), News-Based Indices, Geopolitical Risk Index (GPR), Event Data, GDELT, ACLED.

Introduction

Geopolitical risk refers to the uncertainty and instability stemming from events like conflicts, terrorism, and tensions between nations [Caldara and Iacoviello]. Such risks can significantly impact global markets and economic stability, affecting investment decisions and business operations worldwide. Recent surveys of executives and investors underscore the importance of geopolitical threats, ranking them among the top concerns for businesses and policymakers. In this context, quantifying geopolitical risk has become crucial for anticipating potential crises and responding proactively. **Open-source data** – publicly available information ranging from news articles to social media and satellite imagery – offers a rich resource for measuring these risks in real time. By leveraging open-source intelligence (OSINT) techniques, analysts can cut through vast streams of unstructured information to derive credible, data-driven indicators of geopolitical risk [Bifolchi]. This approach enables evidence-based decision-making and helps organizations build resilience in a volatile global environment.

Defining Geopolitical Risk

Geopolitical risk can be defined as **the risk associated with wars, terrorist acts, and inter-state tensions that disrupt the normal and peaceful course of international relations**. In practical terms, it encompasses both the threat that such adverse events may occur and the potential for escalation

of ongoing conflicts. For example, heightened military frictions, diplomatic standoffs, or major terrorist attacks all contribute to geopolitical risk. These events introduce uncertainty into global affairs, which can **trigger market volatility, disrupt trade and supply chains, and dampen economic activity** [Caldara and Iacoviello]. A notable characteristic of geopolitical risk is that it often materializes unexpectedly and can have far-reaching spillover effects – not just in the directly affected region but across the world due to today's interconnected markets. This broad definition guides the quantification efforts: to measure geopolitical risk, one must track indicators of conflict and tension across various domains (political, military, economic) and gauge their intensity over time.

The Role of Open-Source Data (OSINT)

Open-source data refers to information that is publicly available, as opposed to classified or proprietary intelligence. **Open Source Intelligence (OSINT)** is the practice of collecting and analyzing such information for insights on global events and trends. In modern geopolitics, OSINT has proven indispensable for risk assessment because much of what happens on the world stage leaves a trace in open sources. News outlets, social media platforms, online databases, and even satellite feeds provide real-time signals about evolving situations. By leveraging these sources, analysts can detect emerging conflicts or shifts in public sentiment earlier than traditional intelligence alone might allow. For instance, spikes in certain keywords on Twitter or news sites may indicate rising tensions or unrest in a region. Publicly available satellite imagery can reveal military buildups or infrastructure disruptions. Shipping and flight trackers expose changes in trade routes or troop movements. All these **OSINT inputs, when systematically gathered and processed, help quantify geopolitical risk by providing timely, granular evidence of instability**. An important advantage of open source data is its transparency and breadth: it covers virtually every country and domain, ensuring that even subtle indicators (like local protests or diplomatic disagreements) can be captured. However, making sense of OSINT requires robust methodologies – including automation (e.g. AI algorithms) to sift through big data, and human expertise to verify and contextualize the findings. As one expert noted, OSINT's strength is its "immense large-scale power" to detect patterns, but it should be complemented with human analysis (HUMINT) for depth and accuracy [Castro].

Methods to Quantify Geopolitical Risk Using Open Sources

Quantifying geopolitical risk with open source data involves transforming qualitative information (e.g. news text or event reports) into quantitative indicators or indices. Several approaches have been developed, often combining advanced data analytics with political expertise:

- **News-Based Risk Indices:** One foundational method is to use news media as a barometer of geopolitical tension. A prominent example is the *Geopolitical Risk (GPR) Index* created by Dario Caldara and Matteo Iacoviello, which counts the frequency of newspaper articles discussing geopolitical threats and conflicts. By scanning articles in major international papers for terms related to war, terrorism, and political crises, their index provides a monthly time series of global geopolitical risk. This news-based index reliably spiked around major events – for instance, it surged during the Gulf War (1990-91), after the 9/11 attacks in 2001, during the 2003 Iraq invasion, and amid the 2014 Russia–Ukraine crisis. Such spikes validate that news intensity reflects real-world risk escalations. Furthermore, the GPR Index research showed that **high geopolitical risk often foreshadows economic repercussions**, including declines in industrial output and stock market drops as investors flee to safety [Caldara and Iacoviello]. Building on this approach, other analysts have used news articles in multiple languages and local sources to create more tailored indices. For example, BBVA Research developed a Geopolitical Risk Monitor that leverages a massive news database (GDELT) to track country-specific risk sentiment on a daily basis, using text analytics and machine learning

to gauge the tone of media coverage [Leetaru and Schrodt]. These news-based indicators convert unstructured text into numerical scores, making it possible to compare risk levels over time and across regions.

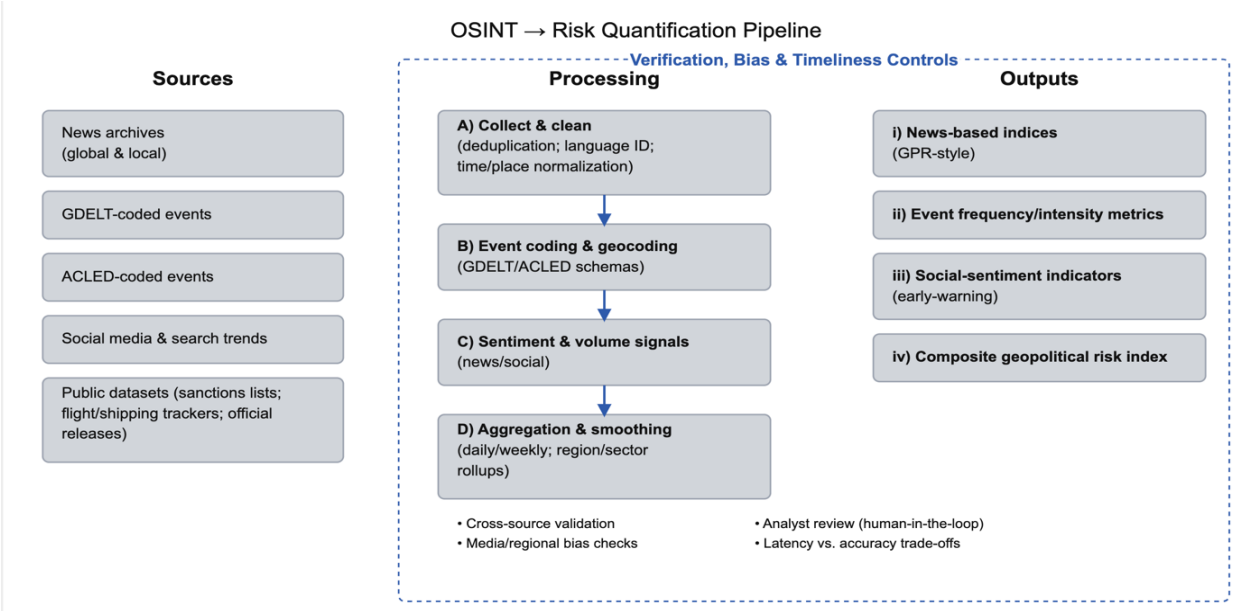
- **Event Databases and Conflict Data:** Another quantitative approach relies on structured databases of geopolitical events derived from open sources. Projects like **GDELT (Global Database of Events, Language, and Tone)** and **ACLED (Armed Conflict Location & Event Data Project)** systematically capture reports of protests, violence, conflicts, and diplomatic events worldwide [Leetaru and Schrodt], [Raleigh et al.]. GDELT, for instance, monitors news from around the globe in over 100 languages, coding every event (e.g. a riot, a peace treaty, a military clash) with its location, actors, and an intensity score. This yields a high-resolution stream of data – updated in near real-time – that can be aggregated into indicators of instability. Similarly, ACLED compiles verified information on political violence and protests from local news and NGO reports, allowing analysts to quantify conflict frequency and spread in various countries [Raleigh et al.]. By counting incidents (and their severity) or summing up risk scores, one can create **metrics such as "number of conflict events per month" or conflict intensity indices** for a given region. These metrics directly measure geopolitical risk in terms of on-the-ground events. Open source event data has been used to generate maps of conflict hotspots, monitor trends in civil unrest, and even assess the effectiveness of ceasefires or peacekeeping missions quantitatively. The advantage of event-based measures is their tangible nature: they record actual occurrences of instability (bombings, battles, protests), which are clear manifestations of geopolitical risk that can be tracked over time.
- **Social Media and Sentiment Analysis:** The explosion of social media provides another open data avenue for risk quantification. Platforms like Twitter, Facebook, and regional forums often serve as **early warning systems** for geopolitical unrest, since local eyewitnesses and citizens post about incidents in real time. By mining social media data, analysts can capture shifts in public sentiment or detect emerging crises before they hit major news outlets. For example, a sudden surge in tweets about troop movements or violent clashes in a particular area might indicate escalating conflict risk. Natural Language Processing (NLP) algorithms are employed to analyze the content of millions of social media posts, classifying whether the tone is fearful, angry, or calm regarding a geopolitical issue. Researchers can quantify this as a sentiment index – a rising proportion of negative or alarmist posts could signal rising risk. Additionally, **trend analysis** on search engine data (like Google Trends for terms related to war or terrorism) can reflect public concern about geopolitical events. Social media-derived indicators have been used in models to predict events such as protests or election-related violence by noting online chatter patterns. The key benefit here is speed and localized insight: open data from social channels often provides granular, on-the-ground perspectives and immediate signals of risk that might not yet be in official reports.
- **Composite Risk Indices (Multi-Source):** Some of the most powerful quantification methods combine multiple types of open source data into a single **composite index**. By integrating economic, military, and political indicators, these indices aim to capture the multifaceted nature of geopolitical risk. A recent example is the *Geopolitical Annual Trade Risk Index (GATRI)* developed by HCSS in 2025, which aggregates diverse open-source metrics to evaluate how global instability affects trade. GATRI incorporates structured data such as the number of new sanctions imposed, changes in military deployments, and counts of trade disruptions, alongside textual measures like diplomatic sentiment (e.g. tone of leaders' speeches) and frequency of high-level state visits [HCSS]. All these components are drawn from public sources – for instance, sanction lists, defense ministry releases, news reports on trade, and international relations databases. By weighting and normalizing the components, the index produces a single score (indexed to a baseline year) that reflects overall geopolitical stability or risk in a given year. An upward move in such an index suggests easing tensions and a safer environment for trade, whereas a downward move indicates mounting geopolitical instability. **Composite indices** provide a holistic view and can be tailored to specific interests; for example, an investor-focused index might combine indicators of expropriation risk, political violence, and currency instability. Many private sector firms (such as risk consultancies and investment companies) have developed their own composite geopolitical risk scores using open data inputs. BlackRock's Geopolitical Risk Indicator (BGRI), for instance, uses text analysis of brokerage reports and news to gauge market attention to

geopolitical threats, translating qualitative concerns into a numeric gauge of risk sentiment [BlackRock]. In all these cases, the transparent use of open data means that the index can be explained and replicated, which builds trust in its findings.

- **AI and Advanced Analytics:** Given the volume of open source data generated daily, artificial intelligence has become a key enabler for quantifying geopolitical risk. Machine learning models can ingest **massive datasets – news feeds, social media streams, satellite images, economic figures – and identify patterns or anomalies** that precede geopolitical events. For example, AI systems might detect an unusual clustering of negative news sentiment, currency fluctuations, and troop movement indicators that together signal a high probability of conflict outbreak. Natural language processing algorithms classify text from news and diplomatic communications to track how narratives and rhetoric are shifting. These techniques allow analysts to move beyond simple counts of articles or events, towards recognizing complex warning signs (e.g. a combination of rising inflation, surging online protests, and military drills could flag an emerging crisis). The use of AI with open data has already shown practical results: there are cases where AI-driven alerts from open source inputs gave companies weeks of advance warning about risks like port closures or new trade tariffs, enabling them to reroute supplies and avoid losses. The **speed and scalability** of automation mean that risk models can update in real-time as new data comes in, providing up-to-the-minute risk assessments. However, AI models must be carefully validated – they are only as good as the data they learn from, and they may struggle with novel situations or deception campaigns. Thus, human expertise remains vital to interpret AI outputs and ensure that the quantified risk signals make sense in context [Mezzi].

Figure 1 summarizes the OSINT→risk-quantification pipeline used in this paper. Sources (news archives, GDELT/ACLED event feeds, social media/search trends, and public datasets) flow through four stages—**A** collect & clean, **B** event coding & geocoding, **C** sentiment & volume signals, **D** aggregation & smoothing—to produce outputs aligned with the methods above: **(i)** news-based indices, **(ii)** event frequency/intensity metrics, **(iii)** social-sentiment indicators, and **(iv)** a composite geopolitical risk index. The dashed governance frame denotes verification, media/regional bias checks, analyst (human-in-the-loop) review, and latency–accuracy trade-offs; the curved arrow indicates iterative refinement of source weighting/filters.

Figure 1.



End-to-end pipeline from open sources to risk indices

Case Examples of Open-Source Risk Quantification

To illustrate these methods, it's helpful to consider a few concrete examples where open source data has been successfully used to measure geopolitical risk:

- **Geopolitical Risk Index (GPR):** This index by Caldara and Iacoviello is a landmark example of quantification using open data (news archives). By scanning hundreds of thousands of newspaper articles dating back to 1900, the researchers quantified geopolitical risk on a consistent scale The GPR index clearly highlights historically tumultuous periods – it peaks during World War I and World War II, reflects the Cold War tensions of the early 1980s, and shows an upward drift in the 21st century as new conflicts and global terrorism concerns have risen. Not only does it track history, but it's updated monthly to signal current risk levels. Notably, in **early 2022 the index registered a sharp jump** coinciding with the Russian invasion of Ukraine, signaling one of the highest risk episodes in recent decades (comparable to post-9/11 levels). Because it is derived from openly available news, the GPR index has been widely adopted in economic research – analysts correlate it with market and macroeconomic data to estimate impacts. For example, studies found that when the GPR index rises, investment and hiring tend to slow and stock market volatility increases [Caldara and Iacoviello]. This demonstrates how a quantitative measure of geopolitical turmoil can be plugged into forecasting models or stress tests for businesses. The GPR index is publicly accessible, allowing anyone to monitor geopolitical risk over time or use it in their own risk assessments.
- **GDELT and Conflict Trackers:** The GDELT database, launched by Kalev Leetaru and Philip Schrodt, showcases the power of automating open source analysis. It continuously scans global news in realtime and applies algorithms to identify thousands of event types – from peace talks to suicide bombings – each coded with actors, locations, and even sentiment scores [Leetaru and Schrodt]. Researchers and institutions have built interactive dashboards on top of GDELT to watch geopolitical risks evolve. For instance, one can visualize conflict intensity in various countries by plotting the number of violent events reported per week. During the Arab Spring (2011), such data showed notable surges in protest events across Middle Eastern countries, quantifying the spread of instability. ACLED provides another case: by manually vetting media and NGO reports, it compiles detailed conflict event datasets in Africa, Asia, and beyond, which are used by NGOs and governments to identify **hotspots of violence and forecast where conflict might flare next** [Raleigh et al.]. A practical example of use was in 2020, when aid organizations used ACLED data to anticipate where violence in the Sahel region of Africa was escalating, allowing them to position humanitarian resources in at-risk areas. These event-based datasets underscore how **open source reporting can be translated into actionable risk metrics** – effectively counting the pulses of global unrest.
- **Corporate Risk Platforms:** Several private companies have developed platforms that quantify geopolitical risk for clients, heavily drawing on open data. For example, **GeoQuant** is a startup that produces high-frequency risk scores for countries by using algorithms on news, social media, and expert data. Their system might rate country X's political stability on a daily basis, alerting investors if the score deteriorates sharply (perhaps due to an unfolding crisis captured in media) [Mezzi]. Similarly, major financial firms have their proprietary models: BlackRock's BGRI mentioned earlier scans broker reports and media to see how much attention is being paid to geopolitical issues, effectively measuring "market concern" about geopolitics as a risk factor. These platforms often integrate with user dashboards, so a multinational corporation can monitor a map of the world showing live risk levels, all derived from open intelligence sources (news reports, government statements, online chatter). One concrete case was during the height of U.S.–China trade tensions: risk analytics firms using open data were able to quantify the likelihood of new tariffs by analyzing sentiment in Chinese state media and U.S. political speeches. When their indices indicated a rising probability of trade conflict, some companies preemptively increased inventories or shifted supply chains – illustrating the value of quantified risk signals. Another example is an AI-powered geopolitical risk platform (as described in a recent pitch by an OSINT firm) that integrates an **event-entity knowledge graph** with a simulation engine. It allows users to see connections between people, organizations, and events in open data, and to run scenario analyses (e.g., how would a new sanction on Country Y ripple through supply chains).

While these corporate tools are often proprietary in their algorithms, they rely on the **backbone of open source inputs** and demonstrate how the data can be leveraged to create tailored, quantifiable risk metrics for decision-makers.

Challenges and Limitations

While open source data has revolutionized geopolitical risk quantification, it comes with several challenges that practitioners must navigate:

- **Data Overload and Noise:** The sheer volume of open source information is overwhelming. Millions of news articles, social posts, and reports are generated daily. Not all of it is relevant to geopolitical risk; the signal-to-noise ratio can be low. Automated systems may pick up false alarms (for example, an inflammatory but unsubstantiated social media rumor) that need to be filtered out. Ensuring data quality is critical – poor or biased inputs can skew the risk metrics. Analysts have noted that **data reliability varies widely across sources**, and algorithms must be tuned to weigh trusted sources more heavily [Mezzi]. Moreover, languages and regional coverage matter: open data might be plentiful for some countries but sparse for others (due to censorship or less media activity), potentially underrepresenting risks in those areas.
- **Verification and Accuracy:** Open source reports are not always accurate or unbiased. They may contain propaganda, misinformation, or errors. This is a limitation because a quantified risk model is only as good as the data feeding it. OSINT analysts therefore spend effort cross-verifying events across multiple sources. For example, a single tweet about a "military coup" could be a false rumor; confirming it via reputable news outlets or official statements is necessary before treating it as a data point. **Skilled human analysts are needed to validate and interpret the results that algorithms produce** [Bifolchi]. As Giuliano Bifolchi explains, OSINT requires critical thinking and expertise to assess each piece of data's relevance and credibility. Automated tools can flag anomalies, but humans must judge whether an apparent spike in risk is real or a glitch (perhaps caused by a flurry of media attention on a minor issue).
- **Bias and Framing in Media Data:** Relying on news sources means that media biases can influence the risk indicators. Certain conflicts might be over-reported by international press (making risk seem higher) while others are under-reported (e.g., crises in remote regions with few journalists). Additionally, open data picks up what is talked about publicly; some geopolitical risks might simmer under the radar with little media coverage until they explode. Quantitative models struggle with this "unknown unknown" problem – they can miss emerging threats that haven't yet surfaced in open discourse. Analysts attempt to mitigate this by broadening the range of sources (including local newspapers, specialized blogs, and so on) and by updating dictionaries and keywords as new issues arise. Nonetheless, users of open-source risk metrics should be aware that **a low score isn't always reassuring** – it might mean either genuinely low risk or simply lack of data visibility.
- **Dynamic and Complex Causality:** Geopolitical risks are complex and do not always follow historical patterns. A purely data-driven model might be blindsided by unprecedented events (for instance, a sudden diplomatic collapse or an unconventional warfare tactic) that don't match prior data. AI models trained on past data have difficulty handling unique shocks or rapid regime changes since those situations present **non-linear, human-driven dynamics**. For example, the outbreak of a war might hinge on a leader's personal decision – something not easily predictable from open data trends. This limitation means that quantification should not be seen as infallible prediction. Instead, it's a support tool that highlights probabilities. Human geopolitical experts are still needed to interpret when an outlier event might occur despite benign indicators. As one analysis noted, algorithmic tools struggle with "unpredictable human decisions, rare events, and data quality issues," reinforcing that expert judgment must complement the models [Mezzi].
- **Integration and Timeliness:** Another challenge is integrating different data types (text, numeric, geospatial) in a meaningful way and doing so fast enough. Geopolitical crises can develop within hours, so risk metrics need to update accordingly. Projects have made progress on real-time dashboards, but some data (like comprehensive conflict databases) are updated with delays due to verification processes. There is often a trade-off between speed and accuracy. Over-reliance on instantaneous social media data

could trigger false alarms, whereas waiting for confirmed reports might mean the risk metric lags behind reality. Finding the right balance is an ongoing effort in the field of open-source risk quantification.

Conclusion

Quantifying geopolitical risk with open source data has evolved from a niche research endeavor into a vital component of modern risk management. By harnessing the **wealth of publicly available information** – from news articles and conflict event logs to tweets and satellite photos – analysts can create dynamic indicators that make an abstract concept like "geopolitical instability" measurable and visible. These indicators and indices empower decision-makers in governments, businesses, and financial markets to monitor the global risk landscape continuously and objectively. The examples of the GPR index, GDELT, ACLED, and composite models show that open data can successfully capture real-world geopolitical upheavals and even provide early warnings before crises fully unfold. Moreover, the practice of quantification forces a clearer understanding of risk drivers: by breaking geopolitical risk into data points (attacks, military buildups, sanctions, sentiment shifts), we better grasp how and why it changes over time.

Looking ahead, the integration of AI and OSINT is likely to further enhance our ability to measure and forecast geopolitical risk. We can expect more sophisticated models that draw on **ever-expanding open datasets – including economic indicators, cyber events, and climate-related stress – to paint a comprehensive picture of global risk**. However, the human element remains key. Analysts and policymakers must interpret these quantitative signals wisely, acknowledging their limitations and cross-checking with qualitative insights. In sum, open source data has provided the tools to quantify geopolitical risk in a way that was not possible a few decades ago. Used properly, these tools offer a powerful advantage: the ability to anticipate and mitigate the impact of geopolitical turmoil through data-driven foresight. In a world where geopolitical shocks can erupt with little warning, such foresight is invaluable for maintaining stability and security across borders.

References

1. Bifolchi, Giuliano. "The Importance and Benefits of Open Source Intelligence (OSINT) in Geopolitical Assessments." *SpecialEurasia*, 17 Apr. 2023, www.specialeurasia.com/2023/04/17/open-source-intelligence-osint/. Accessed 5 Oct. 2025.
2. BlackRock Investment Institute. "Geopolitical Risk Dashboard." *BlackRock*, www.blackrock.com/corporate/insights/blackrock-investment-institute/interactive-charts/geopolitical-risk-dashboard. Accessed 4 Oct. 2025.
3. Caldara, Dario, and Matteo Iacoviello. "Measuring Geopolitical Risk." *American Economic Review*, vol. 112, no. 4, Apr. 2022, doi:10.1257/aer.20191823.
4. Castro, Thales. "Geopolitical Risk Analysis in an Uncertain World: The Case of CogniTrends." *Geopolitical Risk Index*, 5 Sept. 2025, www.geopriskindex.com/world/geopolitical-risk-analysis-in-an-uncertain-world-the-case-of-cognitrends/. Interview. Accessed 6 Oct. 2025.
5. "Geopolitical Risk Forecasting for Investors." *Mezzi Blog*. I was unable to locate a specific article or blog with this exact title. However, several sources discuss this topic, including "AI Tools for Geopolitical Risk Forecasting" on *Lucid.Now* and "Geopolitical Risk Meets AI" from *Georgetown University*. Accessed 7 Oct. 2025.
6. Leetaru, Kalev, and Philip Schrodt. "Global Data on Events, Location and Tone, 1979–2012." *GDELT Project*, 2013, data.gdeltproject.org/documentation/ISA.2013.GDELT.pdf. Accessed 8 Oct. 2025.
7. Raleigh, Clionadh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. "Introducing ACLED: An Armed Conflict Location and Event Dataset." *Journal of Peace Research*, vol. 47, no. 5, 2010, pp. 651–660, www.acleddata.com/about/history. Accessed 8 Oct. 2025.

Submitted on 14.10.2025.
Sent for review on 09.12.2025.