

International Journal of Data Science and Big Data Analytics

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

GDEL T Successor Project: “K-N-GDEL T” and Modelling Geopolitical Risk with the Geopolitical Risk Index

Christoph Kohlhepp^{1*} 

¹Principal Data Scientist, Fulcrumbright, Oxford University, Wellington Square, United Kingdom. E-mail: chris.kohlhepp@mailbox.org

Article Info

Volume 6, Issue 1, May 2026

Received : 11 December 2025

Accepted : 13 April 2026

Published : 25 May 2026

doi: [10.67191/IJDSBDA.6.1.2026.68-74](https://doi.org/10.67191/IJDSBDA.6.1.2026.68-74)

Abstract

In June 2025, GDEL T (www.gdel tproject.org) went silent without notice or explanation. Until that time, GDEL T had been a pillar of the OSINT (Open Source Intelligence) community. Rooted in DARPA research, and leaning on work by Philip Schrod t and Andy Halterman and the legendary TABARI & PETRARCH event coding ecosystem, GDEL T took on the challenge to model the globe—the kind of which was previously reserved for state actors and intelligence agencies. The GDEL T dataset had proven immensely useful. As we enter a geopolitically more turbulent era, we can no longer take GDEL T for granted. At Karmaflows, this inspired us to re-write GDEL T from scratch, an odyssey that we vastly under-appreciated at its onset. We named the project K-N-GDEL T which stands for “Karmaflows is Not GDEL T.” We are happy to announce that 4 months later, we have substantially completed K-N-GDEL T.

Keywords: Geopolitics, Artificial intelligence, Machine learning, Open source intelligence, OSINT, GDEL T, DARPA, CAMEO

© 2026 Christoph Kohlhepp. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

1. K-N-GDEL T Features

- Complete rewrite of GDEL T from scratch.
- Architecture “leans on” GDEL T V2 GKG Knowledge Graph Records.
- NLP and Automated Machine Translation.
- Event coding (dictionary based, CAMEO & PETRARCH2) as well as Large Language Model guided “fuzzy” event coding overcoming GDEL T “bag of words” limitations.

* Corresponding author: Christoph Kohlhepp, Principal Data Scientist, Fulcrumbright, Oxford University, Wellington Square, United Kingdom. E-mail: chris.kohlhepp@mailbox.org

2710-2599/© 2026. Christoph Kohlhepp. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

- Sentiment and objectivity scoring.
- Geolocation coding.
- Theme and topic modelling (both top-down and bottom-up).
- County coding with lateral scale-out on a by-country basis.
- Content moderation to mitigate impact of generative A.I. fakes.
- Model guided curation of domain specific databases for custom analytics.
- Bespoke analytics of qualitative concepts on a “latent” scale.

In June of this year, GDELT (www.gdeltproject.org) went silent without notice or explanation. Until that time, GDELT had been a pillar of the OSINT (Open Source Intelligence) community. Rooted in DARPA research, and leaning on work by Philip Schrodtt and Andy Halterman and the legendary TABARI & PETRARCH event coding ecosystem, GDELT took on the challenge to model the globe – the kind of which was previously reserved for state actors and intelligence agencies. The GDELT dataset had proven immensely useful. Indeed, we demonstrated that the GDELT dataset could be used to successfully predict the intentions of state actors ([Kohlhepp, 2025](#)). It seemed no coincidence then, when Israel attacked Iran on 13 June 2025, thus putting the nuclear balance of the world at stake, that GDELT went offline two days later. The conflict ended on the 25th of June and GDELT quietly resumed services a few days thereafter ([GDELT Outage; Israel-Iran 12 Day War](#)).

This suggests that, as we enter a geopolitically more turbulent era, we can no longer take GDELT for granted. At Karmaflows, this inspired us to re-write GDELT from scratch, an odyssey that we vastly under-appreciated at its onset. We named the project K-N-GDELT which stands for “Karmaflows is Not GDELT”.

We are happy to announce that 4 months later, we have substantially completed K-N-GDELT. We did not aim for backwards compatibility but rather generally followed the Global Knowledge Record and Event coding paradigm of GDELT.

A lot has happened in the Natural Language Processing domain, Generative A.I. and Large Language Models since the inception of GDELT. Some of the design choices in GDELT that were appropriate in 2015, may no longer be appropriate today. GDELT took on a Big Data problem of colossal scale in part because GDELT attempted to track the flow of news events through the media landscape. This meant reliance on shallow parsers and dictionary approaches – aka “Bag of Words.” Dictionary approaches tend to fare poorly in respect of Precision and Recall when compared with Large Language Models and compared with modern Natural Language Processing techniques.

In practical terms, this means GDELT is prone to miss events which are not caught explicitly by way of the CAMEO coding scheme. The opposite was also true. GDELT over-classified events until the PETRARCH2 event coder mandated both a source and target actor. Thus “Bob meets Alice” satisfies the criteria for an event, while “Alice got married” would not – as there is no mention of who Alice got married to. Likewise, labels are sometimes misapplied: “Use of conventional military force” being one we saw over-applied frequently. Over time, GDELT addressed a number of these issues – which suggests that some of the variations in the GDELT dataset aren’t reflective of variations in the input data but instead indicative of changes to the event coder. This makes longitudinal interpretation of the GDELT dataset possible but challenging – without intricate knowledge of the system.

One interesting aspect of the GDELT dataset is the separation of global knowledge records and event records. Does not every global knowledge record represent one or more events? “IBM closed up 3 basis points in yesterday’s trading,” is an interesting event for a financial analyst at Wallstreet, but it is uninteresting to a political analyst using the CAMEO coding system to classify events. CAMEO concerns itself only with the language of “conflict and cooperation”. Modern Natural Language Processing offers “relation extraction” whereby the grammatical predicate of a sentence in Parts of Speech tagging (POS) becomes the event. In our

example sentence, the predicate is “closed up 3 basis points.” It’s what happened to IBM shares the previous trading day. This informs that a conflict-cooperation scale is but one way to dissect a dataset. Indeed, without a context scale all grammatical “predicates” become events. Hence the notion of events only makes sense in the context of a scale to be determined “a priori” (CAMEO).

What we would ultimately like is the ability to quantitatively scale arbitrary, latent concepts in a qualitative context - for instance a “Bear Market vs a Bull Market” in the context of “economics” or perhaps more specifically in the context of “commodity equities.” If we can capture this latent trend (or predict it) before the market digests the information, we would stand to capture substantial windfalls. Or we might wish to scale “hostile intentions vs peaceful intentions” in the context of “geopolitical threats.” It is this facility that the K-N-GDELT platform introduces as an analytical capability. It is difficult to overstate the research significance of this facility as compared to a platform that is fixed to just one single scale that is enshrined in its database schema.

Data architecturally then, K-N-GDELT prefers a uniform treatment of global knowledge records without the need for later cross-matching event records that we found to be very resource intensive in GDELT – owing to the nested level of packing of columnar data in SQL tables that need unpacking in a 3rd generational programming language such as C++ or Python. In GDELT, this manual “unpacking” of columns makes sub setting of data up-front in SQL all but impractical.

We also worried a lot about “authority” and “authenticity.” When GDELT was conceived, generative A.I. did not exist. This means GDELT today faces a lot of unmoderated noise - which we had to solve for K-N-GDELT.

In the universe of K-N-GDELT, we are less interested in modelling how stories traverse the media landscape than we are interested in capturing relevant facts, analysing them and ultimately predicting them – eliminating the need for multiple mentions of the same event over time. That in turn suggests we can avail ourselves of “deeper” learning methods by reducing scale. Big Data, yes. “colossal scale,” no. Instead, we added the ability to laterally scale-out on a by-country basis.

Overall, K-N-GDELT, while offering country and continent scale modelling of “Global Knowledge Records,” looks very different from the GDELT V2 or GDELT V3 record standards. K-N-GDELT performs all the bread-and-butter tasks of GDELT, such as event coding, automated translation, geo-coding, as well as theme & topic modelling (both top-down and bottom-up).

2. K-N-GDELT Analytical Capabilities

Below are some of capabilities of K-N-GDELT that poignantly fit the geopolitical context that gave birth to K-N-GDELT (Figure 1).

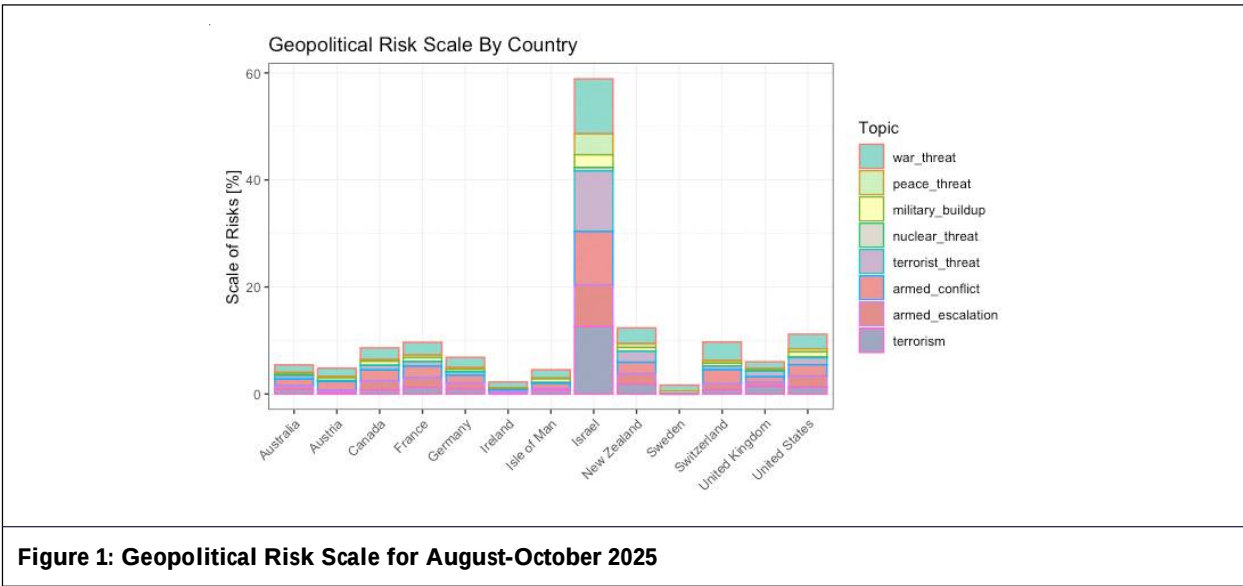
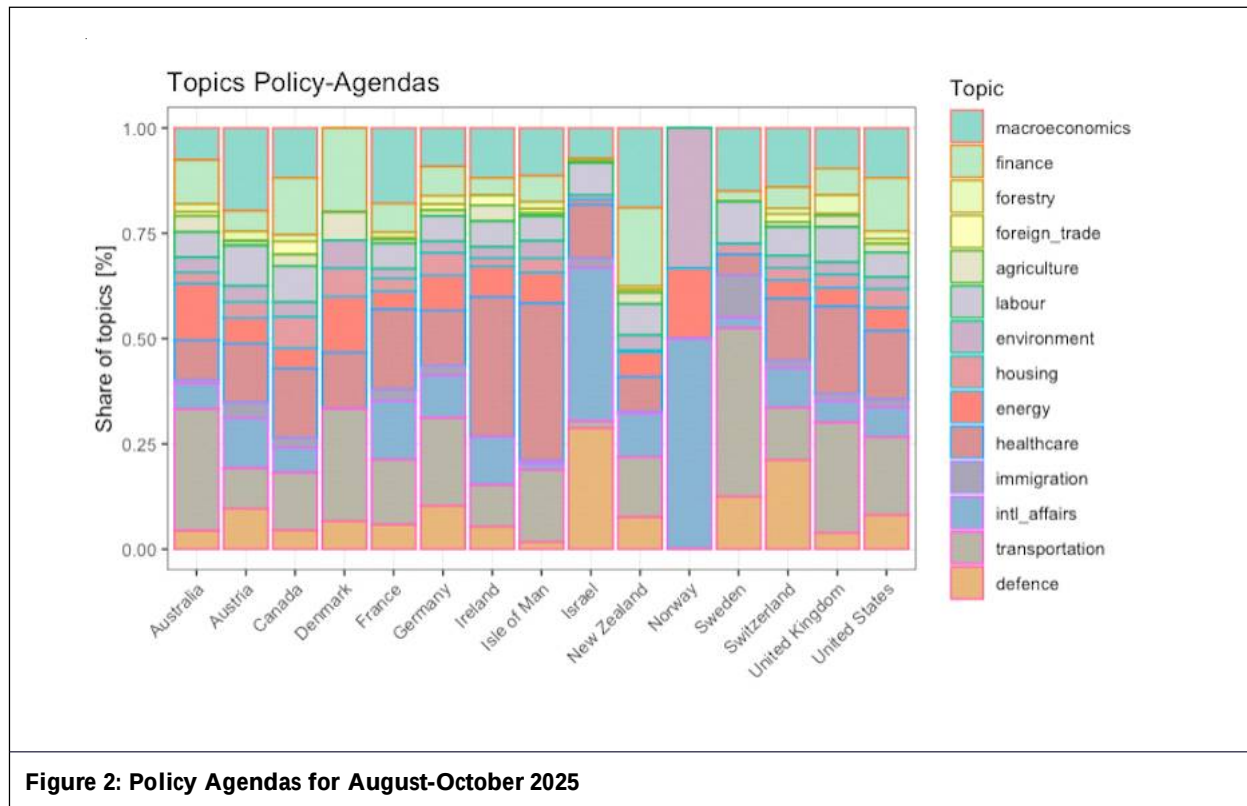
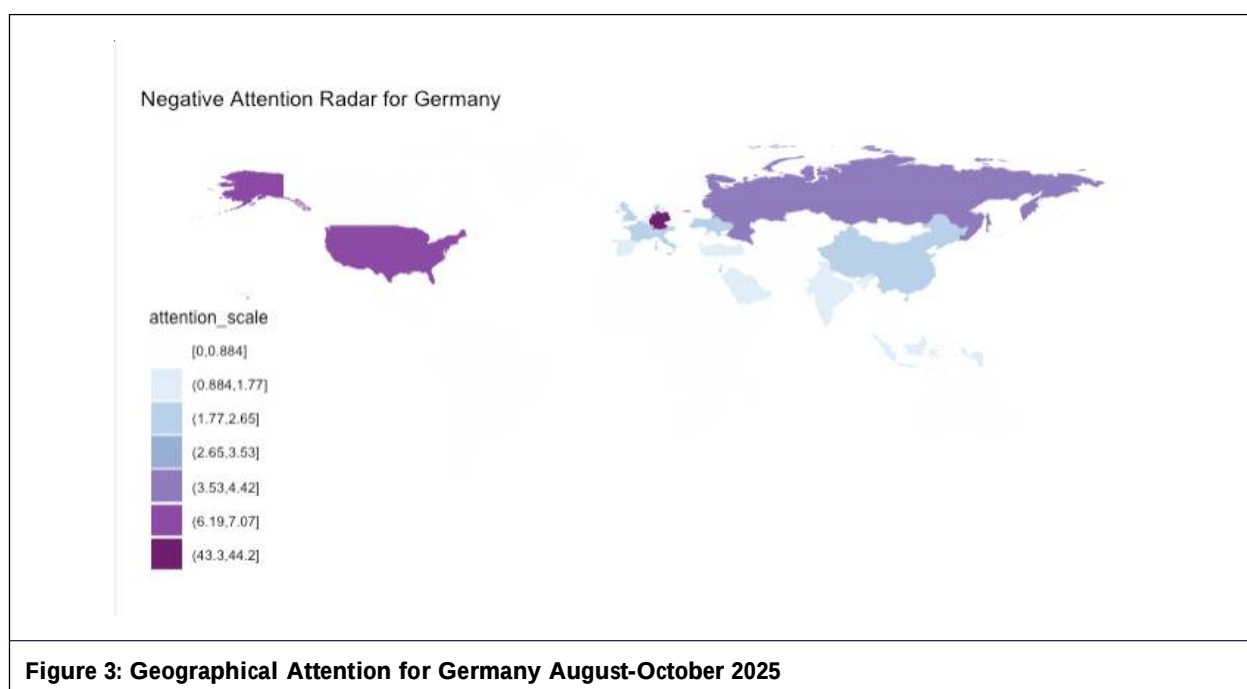


Figure 1: Geopolitical Risk Scale for August-October 2025

What am I seeing? Geopolitical risk for Israel outstrips all else. Armed conflict, terrorism and the treat of a continuation of both dominate the spectrum. This agrees with observation of recent news events, but here we can quantify the risk exposure of various states to geopolitical risk (Figure 2).



What am I seeing? It is often informative to cross check one measure against another. For instance, we observe that for Israel defence and international affairs dominate. It is interesting how relatively unconcerned Australia is about defence in the present geopolitical climate. Switzerland, on the other hand seems more concerned about defence than even the United States. Given the lack of military alliances, this posture on the part of Switzerland is an interesting bellwether. Except for Norway, almost nobody is over concerned about the environment, likewise interesting in an era of global warming (Figure 3).



What am I seeing? This is an interesting facility in K-N-GDELT. It allows inspecting the relative sentiment accorded to third countries as seen by a principal country, here Germany. Filtering can be used to identify positive sentiment, negative sentiment, both in aggregate or on any other variables in the dataset. Clearly, Germany is most “concerned” about relations with the United States, Russia and China (Figure 4).

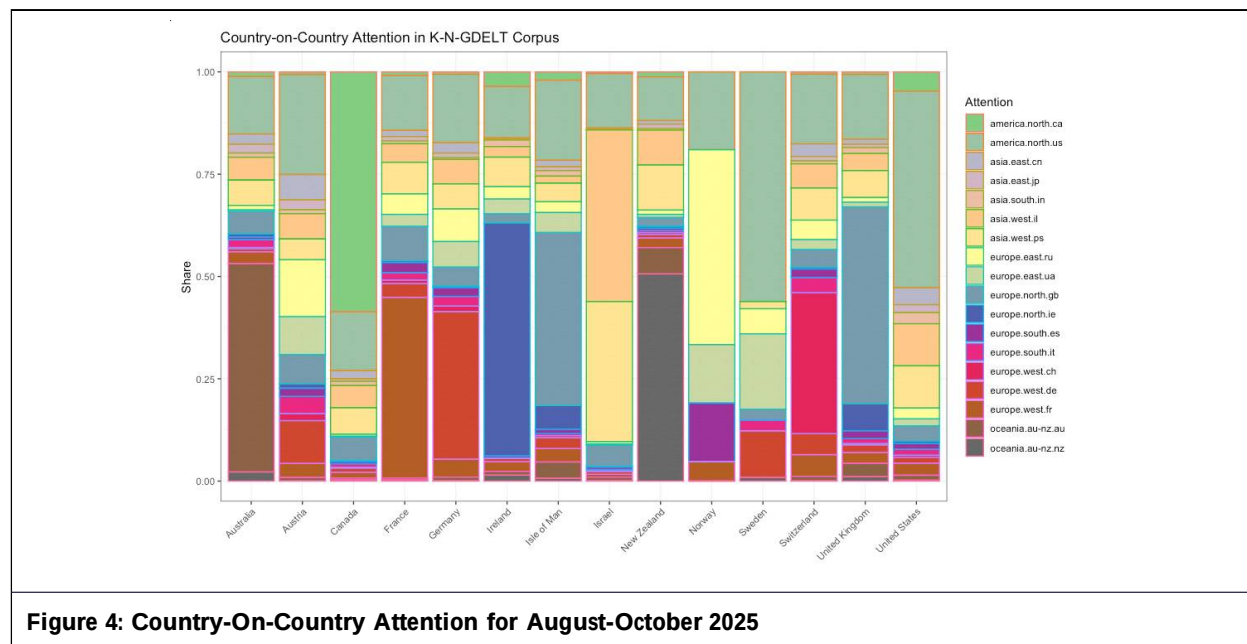


Figure 4: Country-On-Country Attention for August-October 2025

What am I seeing? The country-on-country attention matrix is similar to the geographic attention radar. What stands out is how each country is mostly concerned with itself and direct neighbours. We can also compare how, for instance, various countries are concerned about specific third countries simply by comparing the relevant sizes of coded colour bag segments.

Quantitatively scaling arbitrary, latent concepts in a qualitative context is perhaps the most interesting facility on K-N-GDELT

Integrated in K-N-GDELT is the “Geopolitical Threat Index.” The Geopolitical Threat Index is validated against historical records (Trubowitz, 2021).

As with any index, interpretation is relative to a benchmark. What we usually want to ask is: what stands out? Below we investigate the scale of “Geopolitical Threats” in the context of “Russia.” This produces the follow topic model atlas (Figure 5).

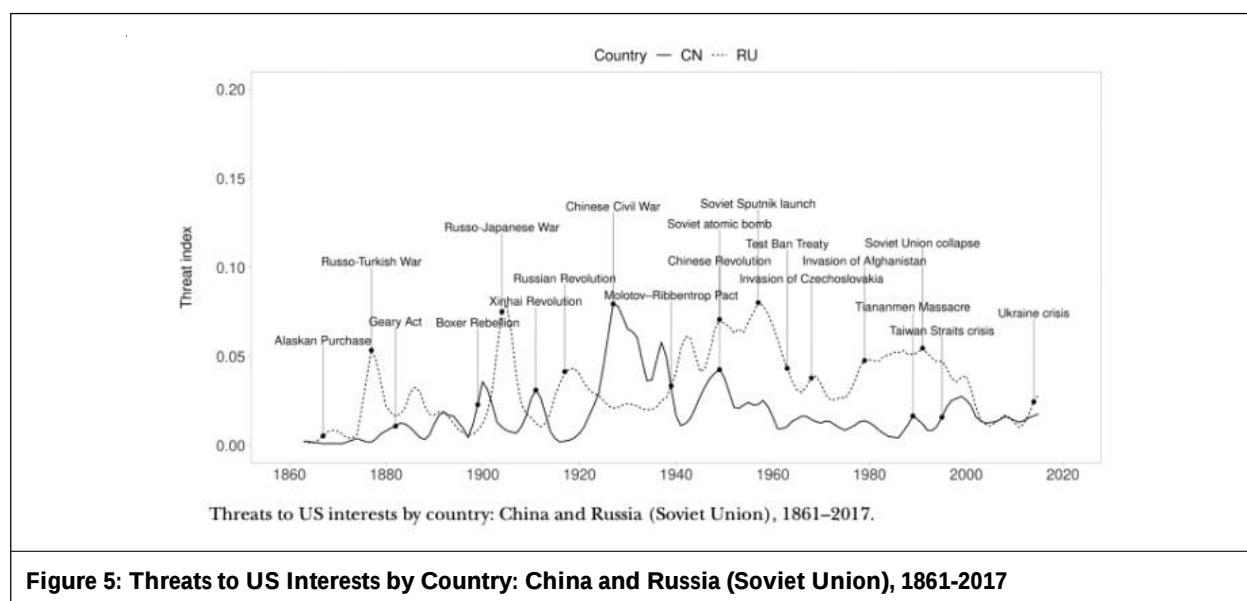
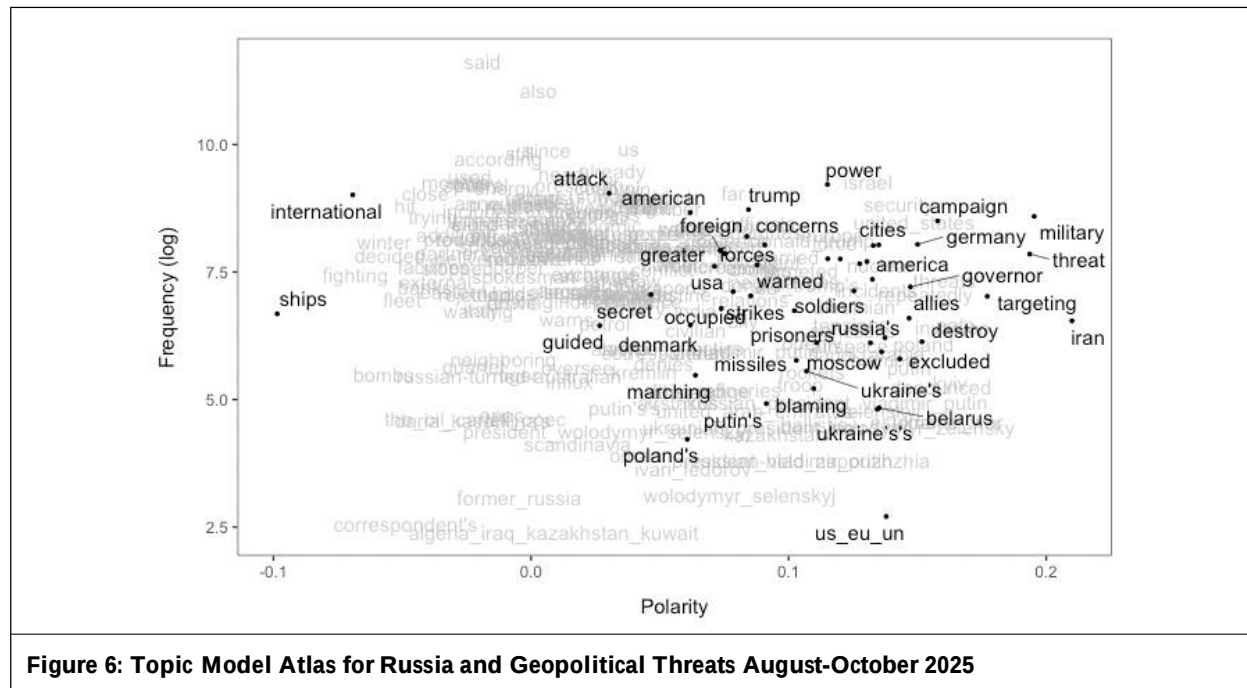
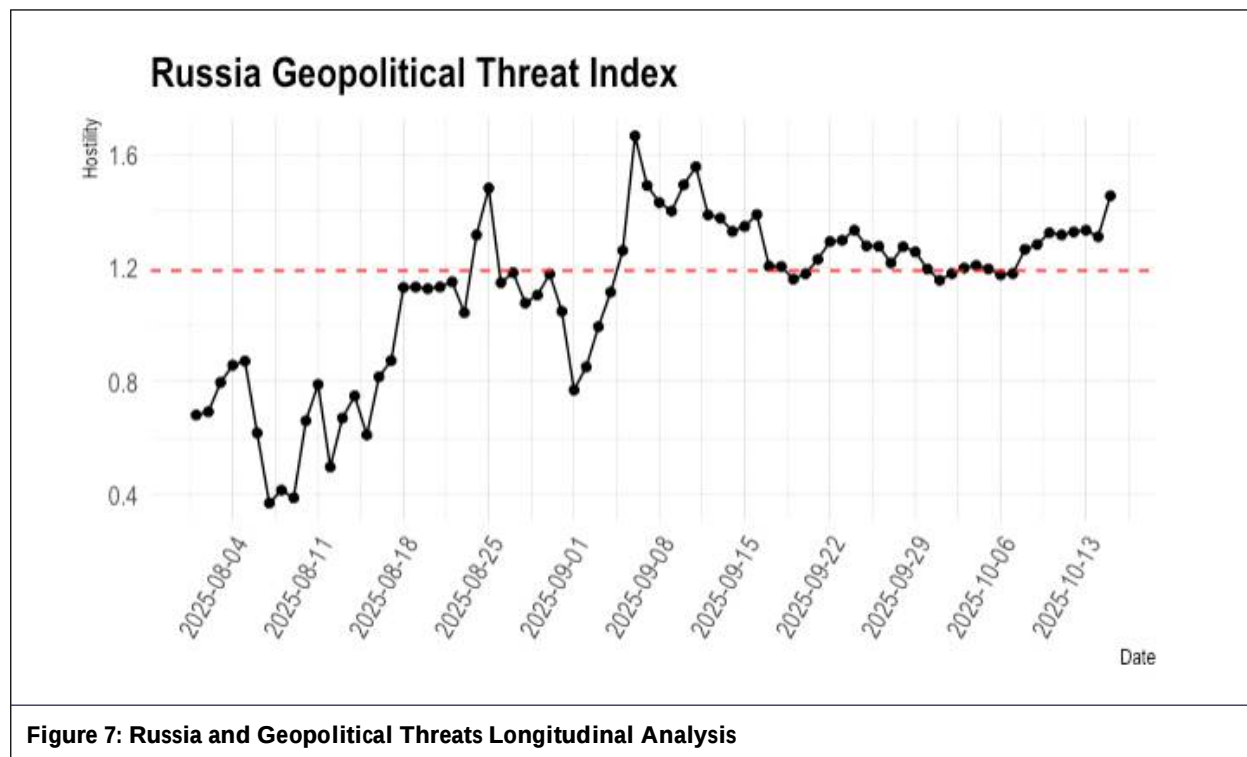


Figure 5: Threats to US Interests by Country: China and Russia (Soviet Union), 1861-2017



We also obtain a longitudinal analysis of the concept of geopolitical threats in the context of Russia for the period of August-October 2025 (Figure 6).

1.2 Marks the “mean” Geopolitical Threat Index in the context of Russia for the period of August-October 2025 (Figure 7). What would be of interest to an analyst is what events and knowledge records contribute to the outliers suggestive of escalation. K-N-GDELT furnishes this analysis together with links to the source articles.



What am I seeing? We observe news sources commonly in English, French & German but extending to languages such as Hebrew. Annotated at “valence” (sentiment) and the Global Threat Index (GTI) rating relative to the 70th percentile of catalogued threats (Figure 8). Top concerns are centred upon the construction of a European drone wall, general mobilisation, hyper sonics and the direct involvement of China in the Ukraine conflict on the side of Russia as indicated by German Military Intelligence.

Leading Geopolitical Threats in the context of Russia with 70th %ile GTI > 1.81 Corpus Mean GTI: 1.19							
uuid	publish_date	title	title_translated	valence	GTI	tax_ru	url
04b1ef55-ee87-45be-b5a2-af88ad0c237f	2025-10-04	Will Europe's 'drone wall' really defend us?		-1.0000000	3.736365	5.998206	http://thesundaytimes.co.uk/wo/late-but-theyre-warning-the-we
8903977b-be53-4337-b9d5-d652c466084b	2025-10-05	Drohnen und Hyperschallraketen: Russland greift Ukraine massiv an	Drones and hyperschallrakets: Russia attacks Ukraine massively	-0.6842105	1.843931	5.681790	https://www.focus.de/politik/be/affenruhe-und-druck-auf-puti-24a02adfd62d.html
bbda7dc9-00ff-4aa9-9a47-8f3de82af7e6	2025-10-02	L'Ukraine appelle l'Europe à se mobiliser pour contrer l'"escalade" de la Russie	Ukraine calls on Europe to mobilize to counter Russia's "escalation"	-0.8888889	2.071861	5.657732	https://www.notretemps.com/de/a-se-mobiliser-pour-contrer-l-es
8315c444-fc2e-4dc8-a683-b5b91268c498	2025-10-02	Sorge um Sicherheit der ukrainischen Atomanlagen	Concern for the security of the Ukrainian nuclear facilities	-1.0769231	2.731991	5.571600	http://ksta.de/politik/dpa-politik/ukrainischen-atomanlagen-112/
32a04571-f28d-4dc2-ae04-4c310cc16912	2025-10-04	Geheimdienst: China versorgt Russland mit Satellitendaten für Angriffe in Ukraine	Secret service: China provides Russia with satellite data for attacks in Ukraine	-0.3684211	2.032271	5.445182	https://www.focus.de/politik/be/affenruhe-und-druck-auf-puti-afe7079af31f.html
7cc36ddb-a2f6-4e0f-a5ae-d7009e5a2c2a	2025-10-04	Person killed in drone attack as Ukraine accuses Russia of targeting passenger trains		-1.5882353	2.035033	5.414900	http://guardian.co.uk/world/202-of-targeting-passenger-trains
2264f098-8451-4b41-a87b-7f8a04595055	2025-09-20	Russia launches large-scale attack on Ukraine, killing three and injuring dozens		-1.0000000	2.174796	5.288509	http://irishexaminer.com/world/
9622aafd-51c1-414f-8d8e-7324e7eb66d8	2025-09-29	حرب التهديدات تتصاعد... "النية الزناد" تعيد الصراع الإسرائيلي الإيراني للمربع الأول	The war of threats is rising .. The "trigger mechanism" restores the Israeli -Iranian conflict for the first box	-0.2083333	3.188160	1.877708	http://eremnews.com/news/world

Figure 8: Leading Geopolitical Threat Events in the Context of Russia

What goes on behind the scenes is the AI guided generation of a custom database relevant to a domain specific context – starting from a general and universal corpus using a machine learning generated “topic model atlas”. K-N-GDELT is capable of curating subject matter databases centred around arbitrary concepts on demand.

References

- CAMEO. <https://eventdata.parusanalytics.com/data.dir/cameo.html>
- GDELT. <https://www.gdeltproject.org/>
- GDELT Outage. https://en.wikipedia.org/wiki/GDELT_Project#Outage
- Israel-Iran 12 Day War. https://en.wikipedia.org/wiki/Iran%E2%80%93Israel_war
- Kohlhepp (2025). Karma Flows-Black Swan Events and Interstate Conflict Artificial Intelligence Global Event Forecasting (June 24, 2025). Available at SSRN: <https://ssrn.com/abstract=5317150> or <http://dx.doi.org/10.2139/ssrn.5317150>
- Trubowitz, Watanabe (2021). The Geopolitical Threat Index: A Text-Based Computational Approach to Identifying Foreign Threats. *International Studies Quarterly*, 65(3), September, 852-865. <https://doi.org/10.1093/isq/sqab029>

Cite this article as: Christoph Kohlhepp (2026). GDELT Successor Project: “K-N-GDELT” and Modelling Geopolitical Risk with the Geopolitical Risk Index. *International Journal of Data Science and Big Data Analytics*, 6(1), 68-74. doi: 10.67191/IJDSBDA.6.1.2026.68-74.