CISS MUNICH WORKING PAPER SERIES
WORKING PAPER 008/2026

LISTENING TO THE DIGITAL PULSE OF DISSENT: HOW ONLINE ACTIVITY CAN HELP PREDICT PROTEST AND CONFLICT

The CISS Munich Working Papers reflect the views of their respective authors. They do not represent the views of the German Armed Forces or the University of the Bundeswehr Munich. The series is published by the Center for Intelligence and Security Studies (CISS) at the University of the Bundeswehr Munich. The series addresses current, historical and strategic issues in the fields of security and intelligence, military affairs and technology, politics and social change. The series provides insights into developments in these areas, offering analyses and ideas for consideration, and outlining potential actions.

Further publications: https://www.unibw.de/ciss/working-paper-series

ISSN 3053-7800

Summary

Predicting political instability has attracted considerable scholarly and policy attention over the last decade, which in turn has spurred methodological and data innovations. This paper presents one such data innovation: digital trace data in the form of Google Trends and Wikipedia data. It demonstrates how Google Trends and Wikipedia page views help predict the intensity of various forms of violence and how changes to Wikipedia pages help anticipate anti- and pro-government protests in autocracies. Google Trends and Wikipedia data are updated in real time and provide global coverage, characteristics many other commonly used data sources struggle with. In addition, the ways in which such data sources can be used to predict political instability are far from exhausted. They are thus a valuable addition to the forecaster's toolkit and provide avenues for future research on different spatial and temporal levels.

Christian Oswald

Political unrest has always carried a certain degree of surprise. Even in an age of sophisticated intelligence systems and constant media coverage, mass mobilizations often seem to erupt suddenly without prior warning. Examples such as the Arab Spring or protests in Iran following Mahsa Amini's death are illustrative. Traditional structural indicators often used in early warning systems may help anticipate which countries or regions are more at risk, but they tend to be of limited help to anticipate the timing when political unrest occurs. Recent research suggests that the first signals of discontent may already be visible online, long before demonstrators reach the streets or the number of casualties increases.

A growing body of work has begun to explore how digital trace data, such as internet search data, can be seen as early indicators of emerging political instability. These sources capture what people collectively pay attention to and debate, providing a near real-time measure of the social and political issues that feature in public (or private) debate. They also offer something that traditional indicators cannot: global coverage, fine-grained temporal resolution, and near-instant updating schedules. Such data characteristics provide ideal conditions for early warning systems.

The Predictive Turn in Conflict and Protest Studies

Forecasting political violence has become one of the more active areas in contemporary political science research. Early models, built largely on structural variables such as GDP, regime type or ethnic composition, improved our understanding of long-term risks but offered limited insight into the short-term dynamics of unrest. The overarching quest is the search for variables which can help anticipate the timing of unrest better. As a result, we have observed a shift from data on the annual level with lagged updating schedules to variables on more disaggregated temporal levels with more regular releases which could have the ability to reflect sudden escalations in tension.

Recent advances in machine learning algorithms and data innovations have enabled forecasts at monthly, weekly, and even daily levels. Related research in turn investigated more when conflicts might intensify and how societies might mobilize. Digital trace data complemented slower-moving socioeconomic indicators and helped improve these efforts. In the absence of improving the quality of existing data sources, data innovations are, together with algorithmic innovations, the most promising paths to improve conflict prediction efforts.

Digital Traces as Political Thermometer

Among the more promising of these digital indicators are Wikipedia activity and Google search trends. Both platforms document what large numbers of people search for, read, and write in real time. Because they are global, multilingual and publicly available, they have the potential to provide an unusually open and transparent view of shifting political attention over time.

Research has found that Wikipedia editing patterns can reflect patterns of public (and private) debate within societies.[1] When politically sensitive issues, such as human rights, elections or censorship, become contested in society, related pages tend to attract editing and discussion. The patterns can be thought of as an online barometer of controversy, where peaks in editing activity coincide with, or precede, protest activity on streets and squares.

In a similar vein, upticks in Google search trends and Wikipedia page views for a particular country may indicate increases in political instability or conflict.[2] The frequency and intensity of searches and page visits rise sharply with international interest in a developing situation. Such collective behavior can be thought of as a digital trace of escalating attention. The signals derived can be valuable indicators at relatively low cost for analysts and researchers in their quest to anticipate violence or large-scale unrest.

I presented this new and promising approach in two recent papers. One focuses on protest dynamics in authoritarian regimes, the other on the intensity of armed violence worldwide. Both draw on openly available data using open-source software and tree-based machine learning algorithms to test whether online attention can predict real-world instability.

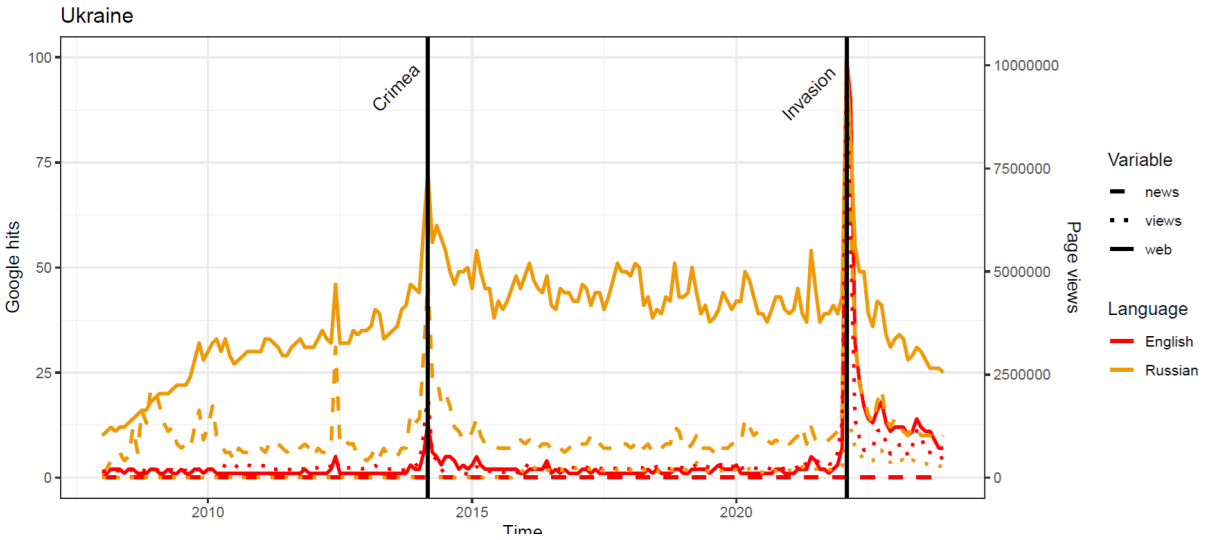

Figure 1: Illustration of Google Trends and Wikipedia page views.

[1] Oswald, Christian. 2025b. "Listening to the wind of change: Predicting protest dynamics in autocracies using Wikipedia page changes." Preprint.

[2] Oswald, Christian; Ohrenhofer, Daniel. 2022. "Click, click boom: Using Wikipedia Data to predict changes in battle-related deaths." *International Interactions* 48(4): 678-696.

Oswald, Christian. 2025a. "I still haven't found what I'm looking for: Predicting security-related incidents and conflict fatalities with Google Trends and Wikipedia data." *Journal of Conflict Resolution*: online first.

Predicting Conflict Intensity globally

This study focused on the number of security-related incidents and conflict fatalities. Previous research was mostly confined to one type of outcome, for example battle-related deaths from the Uppsala Conflict Data Program (UCDP). This paper used different measures of conflict intensity from different datasets: the number of battle-related deaths, and fatalities from violence between armed nonstate actors and violence against civilians from UCDP[3], and the number of security-related incidents from ACLED[4]. Furthermore, the paper predicted these four outcomes both globally on the country-month and for Africa on the province-month level.[5]

Figure 1 illustrates the proposed mechanism and shows Google Trends and Wikipedia page views for Ukraine in English and Russian. Other languages part of the study are Spanish, French, German, Portuguese, and Chinese. One can observe clear upticks in the runup to the invasion of Crimea and the full-scale invasion of Ukraine, which lends initial support for the proposition that international interest in a country increases if there are security-related developments.

Variable	Mean	Median	Standard Dev.	Minimum	Maximum
Views_en	120,138.800	98,335	95,918.690	0	2,516,658
Views_es	21,464.310	15,219	26,626.540	0	989,476
Views_fr	25,130.120	17,800.5	25,199.860	0	526,039
Views_de	22,032.680	17,940.5	20,350.780	0	621,279
Views_pt	10,257.420	6,971.5	13,208.910	0	390,153
Views_ru	10,726.680	6,316.5	17,711.360	0	652,373
Views_zh	4,940.487	3,375	6,793.467	0	347,140
Hits_en_web	31.722	30	19.531	3	100
Hits_es_web	23.229	19	20.110	0	100
Hits_fr_web	25.490	23	18.629	0	100
Hits_de_web	27.213	23.5	19.955	0	100
Hits_pt_web	20.175	14	20.450	0	100
Hits_ru_web	17.213	11	18.945	0	100
Hits_zh_web	7.512	3	12.903	0	100

Table 1: Summary statistics of Wikipedia page views and Google web searches.

[3] Sundberg, Ralph; Melander, Erik. 2013. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research* 50(4): 523-532.

[4] Raleigh, Clionadh; Linke, Andrew; Hegre, Havard; Karlsen, Joakim. "Introducing ACLED: An Armed Conflict Location and Event Dataset." *Journal of Peace Research* 47(5): 651-660.

[5] Oswald, Christian. 2025a. "I still haven't found what I'm looking for: Predicting security-related incidents and conflict fatalities with Google Trends and Wikipedia data." *Journal of Conflict Resolution*: online first.

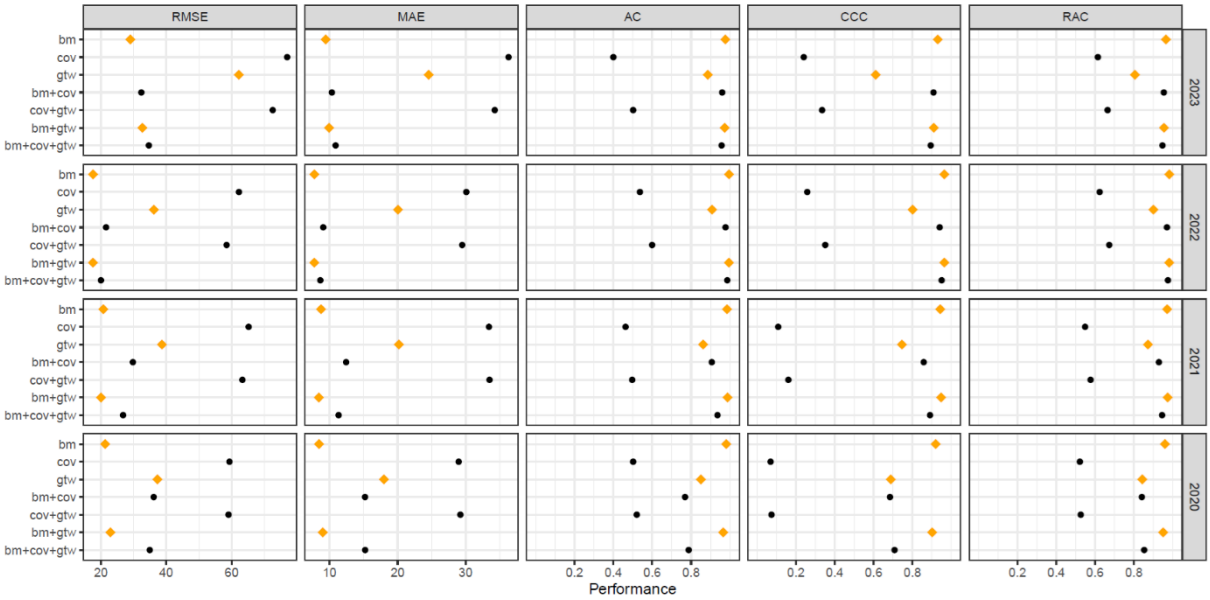

Figure 2: Out-of-sample performance (country-level).

Table 1 shows descriptive statistics for Wikipedia page views and Google web search trends. Google news search trends were also used in this study whereas image and YouTube searches were not utilized. The English language Wikipedia is the most used and frequented by a margin, indicated by the large mean and median values compared to other countries. The Chinese language version, in contrast, is the least actively used one among the seven languages under consideration, while the French, Spanish and German versions are relatively on par and the Portuguese and Russian version are likewise similar but on a lower activity level. Notice that minimum and maximum values vary for Wikipedia page views whereas they range between 0 and 100 for Google Trends. Google does not provide raw data but standardized samples such that trends always show relative rather than absolute interest. However, mean and median values still indicate that the English, Spanish, French, and German searches have more variation than the Portuguese, Russian, and Chinese searches. The variation between languages is used to predict conflict intensity.

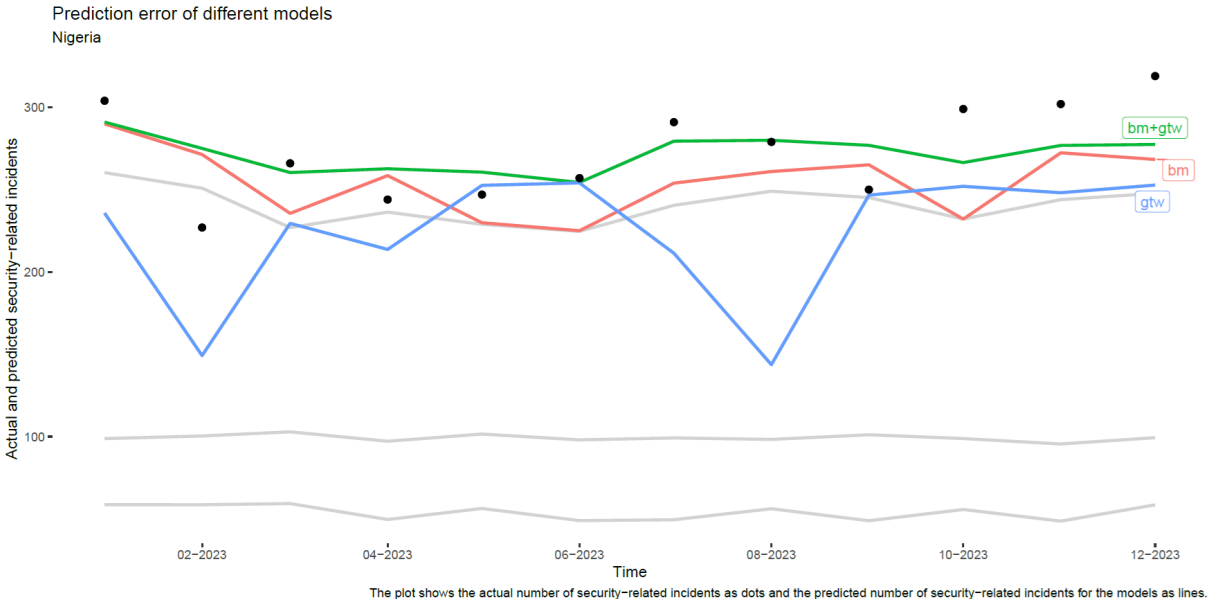

Figure 3: Out-of-sample (Nigeria 2023).

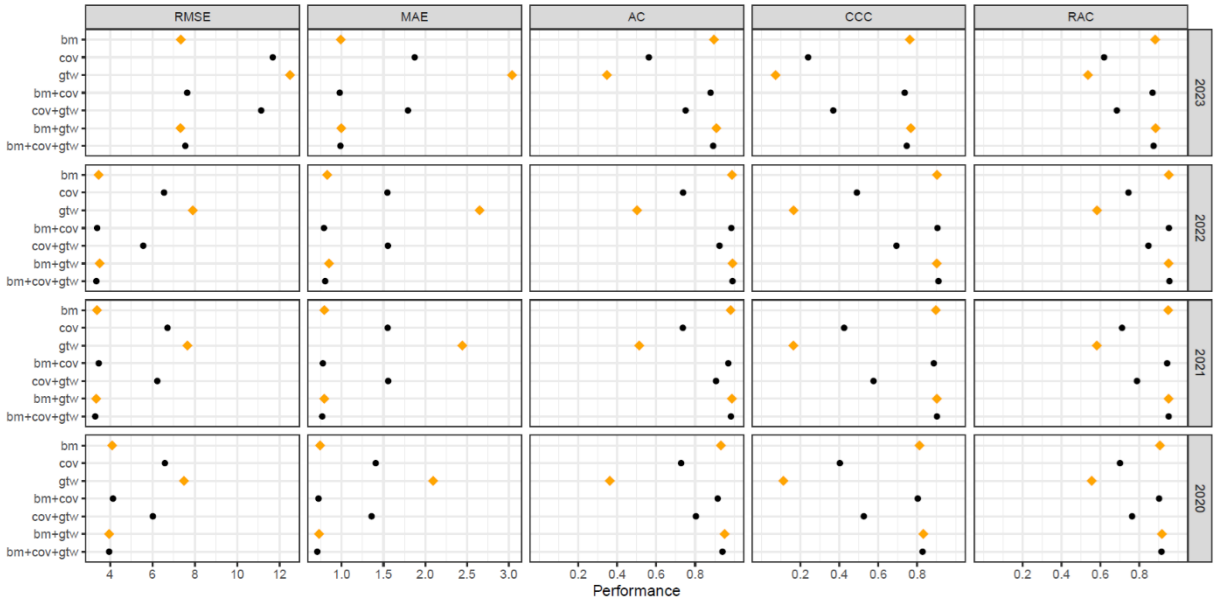

Figure 4: Out-of-sample performance (province-level).

Figure 2 shows comparative results on the country-month level in Africa for four out-of-sample years for the number of security-related incidents. For RMSE and MAE, smaller values indicate better model performance, whereas values closer to 1 indicate better model performance for the remaining metrics. Using both Google Trends and Wikipedia page views in seven major languages does quite well on its own (gtw) but shows its real potential together with information about past violence (bm+gtw). The addition of "salience variables" thus improved the performance of traditional forecasting models at the country level and shows that global online attention can help predict the intensity of various forms of armed conflict. Of note, the approach is also doing well for the different types of fatalities mentioned above, including nonstate armed conflicts, the rarest form.

Figure 3 shows further evidence on the individual country-level. Using the 2023 held-out data for Nigeria, the plot illustrates the predictions of seven models. The Google Trends and Wikipedia (gtw) model and the past violence (bm) model, as well as their combination, are of particular interest. Two issues are of particular interest in this plot. First, Google Trends and Wikipedia seem to do well to pick up general upward and downward trends. Second, the combination of these data sources with past violence makes predictions less conservative. Taken together, the new data source showcases a promising addition to the conflict forecaster's toolkit.

Figure 4 shows evidence on the subnational province-month level. About 18% of all observations saw one or more security-related incidents, compared to 62% of observations on the country-level for Africa and 30% globally. The prediction task is thus more challenging. Notice first the comparatively small RMSE and MAE values, especially compared to the country-

level for Africa shown in Figure 2. The remaining metrics likewise indicate that the combination of Google Trends and Wikipedia page views with information about past violence tends to outperform all other models. Similar patterns to Figure 3 can be observed for individual provinces such as Borno state in Nigeria. In sum, we have evidence that the addition of this novel and dynamic data source is a promising avenue for predicting conflict intensity.

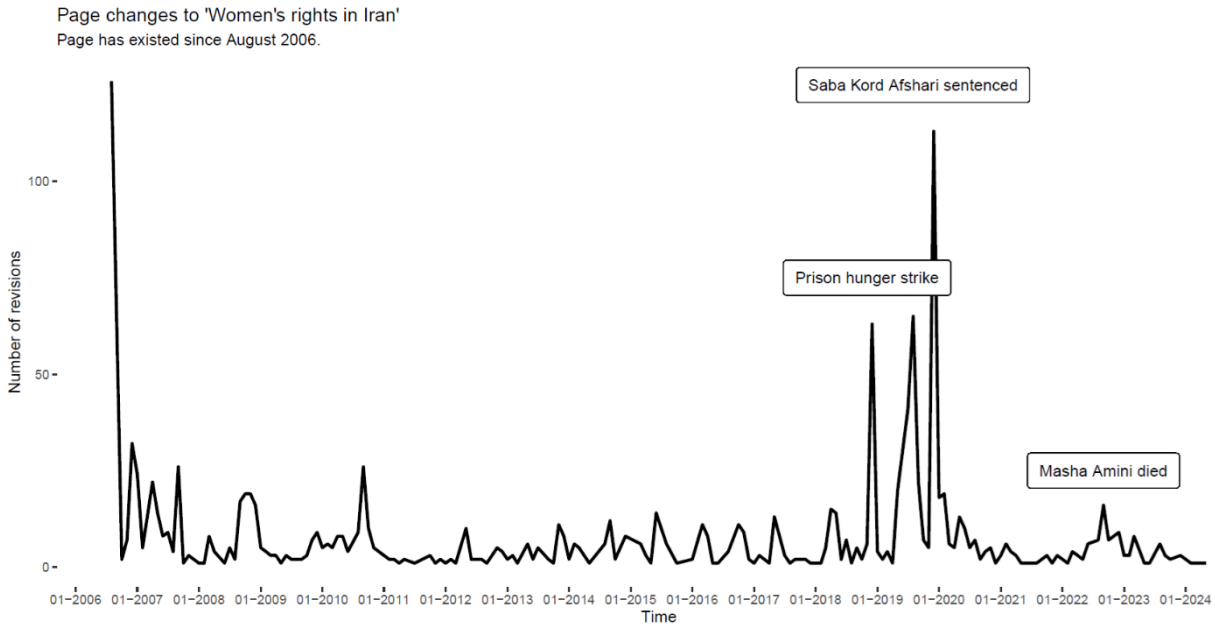

Figure 5: Illustration of Wikipedia page changes.

Predicting Protest Dynamics in Autocracies

Turning to predicting nonviolent conflict, previous research faced a trade-off between low geographical coverage, limited to one or just a few countries, and large amounts of fine-grained "big data" on one side, and wide geographical coverage at sparse temporal resolution, often the country-year, on the other. As a result, predicting nonviolent conflict such as protests is still uncommon and has not seen developments similar to predicting armed conflict, such as global and regional forecasting models at national or subnational units of analysis below the annual level of aggregation.[6] Using changes to Wikipedia pages strikes a balance between wide geographical at low temporal resolutions. Changes to Wikipedia pages covering politically charged topics, such as women's rights, media freedom, or the political system, were used to predict monthly protest activity, both incidences (binary) and numbers (continuous), in 93 autocracies. However, the data would allow to go to lower temporal resolutions such as weeks or days.

Figure 5 illustrates the proposed mechanism using the example of women's rights in Iran. The demonstrations which rocked Iran during this time were widespread and multi-faceted, but women's rights were one integral part of the protest movement's motivation. We can observe spikes coinciding with societal developments, such as hunger strikes, sentencings, or deaths of activists related to women's rights. Societal debates offline can be reflected in real time online. A build-up of developments online can further precede offline activities.

48 country-specific Wikipedia pages that carry a certain debate or controversy potential and may contribute to societal tensions were used, ranging from central banking to agriculture to languages to religion. Both anti- and pro-government mass mobilization events in autocracies were predicted. Previous research has shown that autocrats use pro-government protests to mitigate threats to their survival in power and legitimacy.[7] Therefore, testing the predictive performance of Wikipedia page changes on both anti- and pro-government

[6] Pinckney, Jonathan; Babak RezaeeDaryaKenari. 2022. "When the levee breaks: A forecasting model of violent and nonviolent dissent." *International Interactions* 48(5): 997-1026.

[7] Hellmeier, Sebastian; Nils B. Weidmann. 2020. "Pulling the strings? The strategic use of pro-government mobilization in authoritarian regimes." *Comparative Political Studies* 53(1): 71-108.

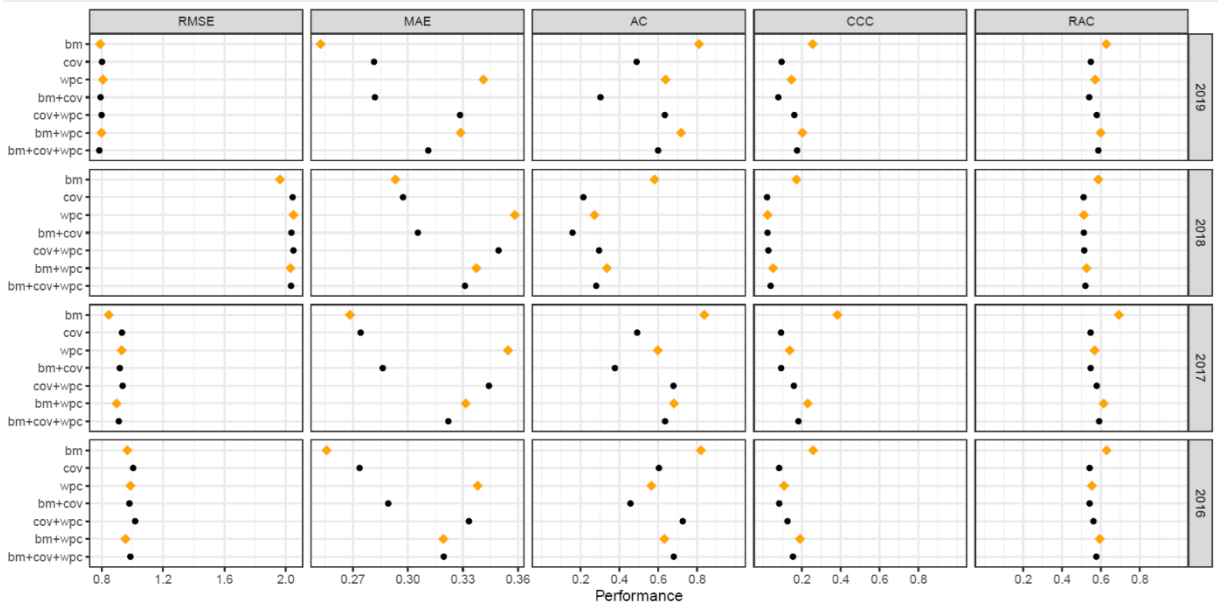

Figure 7: Out-of-sample performance (number of pro-government protests).

protests constitutes a tougher test of the proposed mechanism. If page changes reflect controversy in which different sides and groups participate, they should help predict protests by both sides rather than only anti-government protests. Pro-government protests are also much rarer than anti-government protests, occurring in only 5-7% compared to 25-30% of observations.

Table 2 shows descriptive statistics for revisions of selected Wikipedia pages. Most pages do not get changed very regularly, indicated by median values of 0 or 1. Pages such as censorship, the law or parliament in a country have maximum numbers of changes of around 800 in a given month while the mean number is relatively low. Thus, for prolonged periods of time there can be little movement on these pages, and some months see high numbers of changes. In contrast, ethnic groups pages have large mean values and a median value of 1 whereas the maximum of around 300 is comparatively small, indicating that these pages tend to be changed more frequently. Thus, there is initial support for the mechanism that movement on Wikipedia pages corresponds with real-life changes rather than permanent editorial changes.

The study sought to predict both protest incidences and numbers. We focus on predicting the number of events. Figure 6 shows out-of-sample performance metrics for the monthly number of anti-government protests for four held-out samples. RMSE and MAE indicate that all models struggled more with the 2019 held-out data than with the other samples. Metrics

Variable	Mean	Median	Standard Dev.	Minimum	Maximum
Human rights	1.462	0	5.490	0	262
Womens rights	1.403	0	8.960	0	274
Censorship	0.912	0	7.244	0	787
Ethnic groups	4.897	1	13.489	0	298
Religion	2.206	0	7.919	0	308
Law	1.953	0	18.877	0	823
Parliament	1.460	0	11.051	0	768
Crime	0.673	0	7.332	0	341

Table 2: Summary statistics for selected Wikipedia page changes.

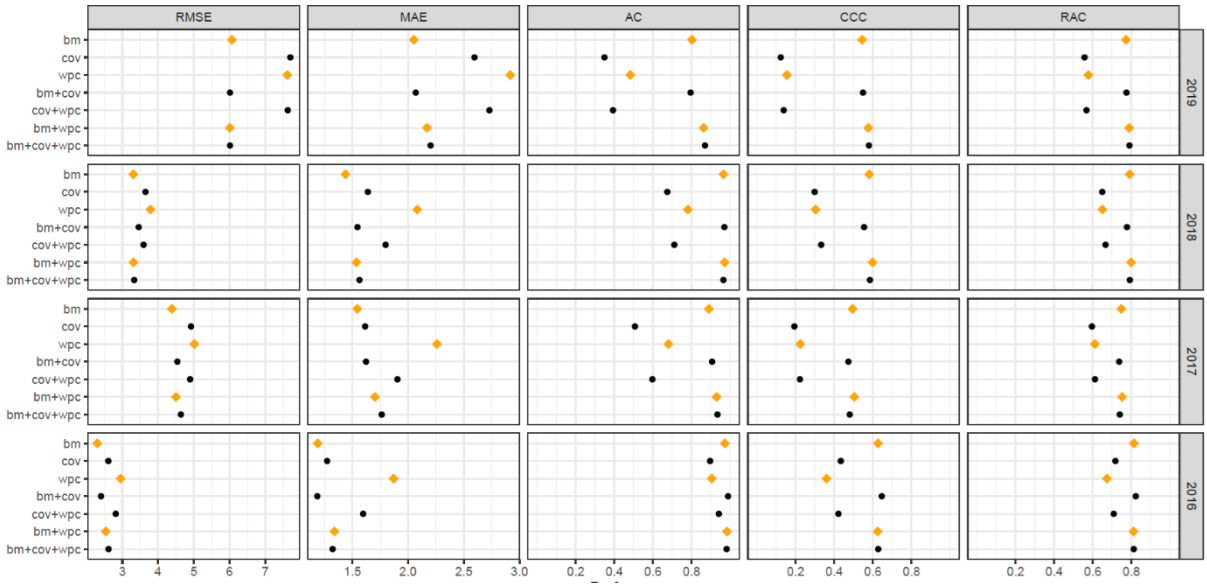

Figure 6: Out-of-sample performance (number of anti-government protests).

which measure general agreement and accuracy, such as CCC and RAC, indicate more persistent performance across held-out samples. Similar to Google Trends and Wikipedia page views predicting conflict intensity, the Wikipedia page changes model does fairly well on its own and is in combination with past protest behavior often the most performant model across performance metrics. It should be noted that models predicting monthly protest incidences reached performance values similar to conflict prediction models, indicating decent performance as well.

Figure 7 shows results for predicting the number of pro-government protests in autocracies. Recall that this is a stress test for the proposed mechanism. The mechanism suggests that there is a struggle between two more opposing sides, seeking to dominate the narrative about unfolding events relating to, for example, women's rights in a country. If this mechanism exists, changes to Wikipedia pages as predictors should be able to predict both anti- and pro-government protests well. Pro-government protests in autocracies are used strategically, especially in circumstances where threats to stability and political survival are anticipated. Notice first that RMSE and MAE values are much smaller compared to models predicting anti-government protests. Notice further that the remaining metrics show similar values compared to anti-government protests, except for CCC. The strong imbalance in the data, i.e. the fact that pro-government protests are much rarer than anti-government protests, explains this difference. Still, the results are encouraging. Similar to anti-government protest incidence, predicting monthly pro-government incidences showed promising results, although the areas under the precision-recall curve are lower in comparison, which is to be expected due to the stronger data imbalance.

In conclusion, incorporating Wikipedia page changes improved predictive accuracy in combination with past protest behavior for many out-of-sample windows and performance metrics. Models using solely page changes likewise showed decent performance. The suggested mechanism seems to be present, since the models performed well for both anti- and pro-government protest incidents and numbers. Thus, digital trace data seem to be a valuable addition to the toolkit of forecasters of armed conflict and mass mobilization.

How Digital Signals Enhance Forecasting

The great value of the digital indicators presented lies in their timeliness and adaptability. Many other indicators which are commonly used, such as GDP or democracy scores, tend to change slowly and are reported infrequently and often on an annual level. Wikipedia and Google data are updated continuously, can be accessed instantly, aggregated at any temporal scale, and integrated with other information sources. They have therefore the potential to bridge the gap between long-term structural risk and short-term situational awareness.

Google Trends and Wikipedia data are relatively transparent and replicable, Wikipedia as an open encyclopedia more so than Google Trends. Any individual can download Wikipedia page views and changes or Google search trends with open-source software. Likewise, all analyses presented in this paper are based on open-source statistical software. As such, the presented data and models fit the call for more transparency and accessibility of operational forecasting systems.[8]

Another advantage is that these sources capture both local and international attention. When societal debates intensify, citizens and observers alike contribute to editing and reading patterns. This dual perspective offers insight into how domestic contention interacts with global awareness.

It is important to reiterate the step forward these data sources represent. Previous research was often restricted to data sources on the annual level and irregular or lagged updating schedules. Especially consistent and timely updates of data sources are critical for operational early warning systems.

Integrating the Digital and the Traditional

While digital trace data offer remarkable immediacy, they are not a substitute for conventional political or socioeconomic indicators. They should best be considered as complements instead. Previous research suggests that structural variables help explain why societies are prone to instability. Digital variables as presented here help capture when those pressures surface in public discourse. Combining the two allows models to detect both long-term vulnerabilities and short-term triggers.

This integrated approach also addresses a persistent challenge in forecasting political violence and instability: models that rely heavily on past conflict events tend to be overly conservative and cautious, predicting the absence of violence until it has already begun. Incorporating real-time digital signals helps make forecasts less conservative and more dynamic and responsive to emerging developments.

Ethical and Practical Considerations

The use of digital behavioral data in social and political science research inevitably raises questions about privacy, interpretation, and bias. In contrast to social media platforms, however, Wikipedia and Google Trends data are aggregated and anonymized, reducing individual-level ethical concerns. These data sources also do not face similar challenges like recent social media platforms such as possible rapid user disengagement or access restrictions, for the time being at least.

Interpretation remains crucial, however. A rise in searches for “Ukraine” might indicate concern about political unrest, but it could equally reflect humanitarian interest or media coverage because of international events unrelated to political stability such as the FIFA World Cup. Analysts must therefore treat online signals as context-dependent indicators rather than definitive measures of dissent, which is one

[8] Rød, Espen Geelmuyden; Tim Gåsste; Håvard Hegre. 2024. “A review and comparison of conflict early warning systems.“ *International Journal of Forecasting* 40(1): 96-112.

reason why the models which combined past activity with digital trace data were among the best performing.

Linguistic and regional biases are of concern as well. Internet access is uneven, which might well mean that, for example, global Google Trends are driven by a particular set of countries. Wikipedia coverage, likewise, is far richer for some countries and topics than others. Further languages of interest, such as Arabic, Urdu or Hindi, do not have as active a Wikipedia as other languages. Efforts to increase language diversity and digital inclusion will therefore enhance the reliability of these tools and expand their set of use cases.

Towards Real-Time Political Analysis

This line of research signals a shift toward continuous political monitoring using openly available data, not unlike the fast-growing subfield of using satellite imagery and remote sensing in conflict research. In combination with event datasets such as ACLED or UCDP, digital salience indicators have the potential to enrich early-warning systems capable of monthly forecasts. Governments, international organizations, and NGOs already take such systems into account to allocate resources and design preventive strategies. Integrating case-tailored online data could help in these endeavors.

The appeal extends beyond forecasting conflict and protests in autocracies. Similar approaches could be applied to study electoral mobilization, misinformation dynamics, or the diffusion of social movements. As computational methods become more widespread across the social sciences, the ability to interpret digital footprints might well become as vital as traditional fieldwork or survey design. Previous research has already demonstrated, e.g., the use of Google Trends to reach difficult-to-survey populations. The potential is far from exhausted.

The Broader Picture

The integration of digital trace data into conflict forecasting represents a significant methodological and conceptual advance at little computational cost with readily available open-source software solutions to obtain and manipulate the data. It demonstrates how everyday online behaviors can generate collective signals about the state of societies. Such signals, when spotted, modeled in context and interpreted carefully, provide invaluable foresight into political dynamics in both more and less open societies.

At the same time, this research reminds us that prediction is not an end in itself. The goal is not to foresee every protest or act of violence, but to build more dynamic and transparent systems that can help decision-makers act before crises spiral out of control or even fully unfold. As digital information flows continue to expand, listening to the world's online pulse may become an essential part of understanding, and perhaps even preventing, the next wave of unrest.

Christian Oswald, Ph.D., is a research associate at the Center for Crisis Early Warning at the University of the Bundeswehr Munich (Germany). He has a background in political science, quantitative methods, and computational social sciences.