Forecasting Armed Conflict Escalation: Adapting the Vision Transformer to a Distribution Prediction Task

Kevin Lin
Dartmouth College

May 28, 2024

Abstract

The literature in forecasting fatalities has increasingly leveraged complex machine learning models as the state-of-the-art. Significant gaps remain in predicting when civil war violence will escalate and de-escalate due to the sparsity of incidences in history [1]. Furthermore, the modern forecasting literature rarely integrates theories of revolution and civil war into predictive modeling. Predicting disparate conflict events and periods of peace are of predominant importance to policy-makers and humanitarian aid groups. Preempting conflict carries multiplicative relevance for implementing an impactful response. We aim to bridge these gaps by proposing a novel theory of conflict initiation and continuation, integrated with novel data augmentation and feature engineering techniques. The forecasting literature has shown that conflicts in Africa tend to have strong spatial and temporal relationships [2]. Multiple machine learning and regression techniques have been applied to account for these attributes. We propose a state-of-the-art Transformer model to exploit these dynamics. Using the PRIO-GRID-month (50km X 50km) as our spatial and temporal unit, we selectively attend to chunks of temporal and spatial data when generating forecasts. Furthermore, we extend the explainability of the forecasting literature by predicting a distribution of fatalities rather than a single value prediction using a Monte Carlo simulation. By refining five model components, the theory of conflict onset and continuation, the model architecture, the data augmentations, the feature engineering, and the outputs, we hope push the state-of-the-art in predictive and interpretability performance of fatality forecasting in Africa.

1 Towards a New Theoretical Model of Violence

Our theoretical model is motivated by the quantitative findings of Hegre, Vesco, and Colaresi in the ViEWS violent escalation prediction competition [3]. Existing models are powerful at predicting ongoing violence based on previous conflict history as patterns of violence are persistent. Likewise, models are powerful at predicting periods of consistent peace as areas with no conflict history are more likely to avoid future violence onset. Furthermore, quantitative deep neural network-based models are typically devoid of theoretical backing through limitations in data collection and labeling [4]. They are motivated by the data-focused forecasting literature rather than the theoretical civil war literature. We therefore separate our model with mechanisms that motivate transitions from peace to conflict (Violent Initiation) and conflict to conflict/peace (Violence Continuation). We ground each mechanism in the literature of civil war onset: grievance, opportunity, greed, and repression. This allows our model to be more interpretable from the policy standpoint and overcomes the explainability problem of understanding cause and effect in deep neural networks [5].

We define two dynamics in a theoretical model of violence in civil war; violence initiation and violence continuation. Violence Initation is the period in which revolutionary actors and groups form with contending viewpoints in opposition to the incumbent state [6]. These actors will engage in varying degrees of detectable forms of overt non-violent resistance with riots, protests, and covert bottom-up resistance [7]. If the state is too weak to suppress overt resistance by revolutionary groups, there is a temporally volatile escalation to confrontation [8]. Once a threshold is crossed from resistance to violent conflict, a mechanism of Violence Continuation defines conflict progression. Revolutionary groups will seek economic resources to fuel the revolutionary conflict, while the state will use a violent coercive apparatus to deter further violence [9], [10].

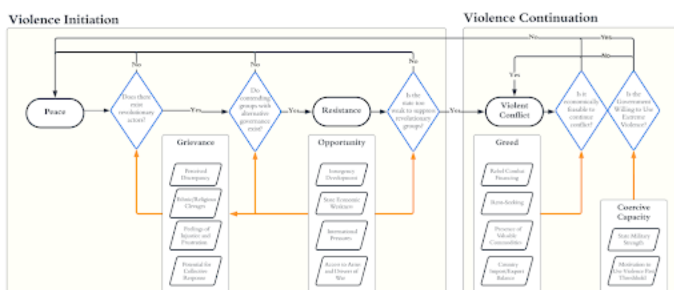

Figure 1. A bilateral model of violence based on grievance, opportunity, greed, and repression [6], [9], [10], [11], [12].

Violence Initiation: Grievance and Opportunity

Fearon and Laitin (2003) contend that civil wars, often perceived as spontaneous eruptions of violence, are more accurately understood as insurgencies: "small, lightly armed bands practicing guerrilla warfare from rural base areas" [12]. The initiation of violence is not solely attributed to grievances from socioeconomic and/or political injustice or sudden political shifts, but rather to a confluence of conditions that foster insurgency [13]. These conditions include state weakness, encompassing financial, organizational, and political fragility, alongside pervasive poverty, challenging terrain, and large populations. Such environments create fertile ground for rebel recruitment, as economic hardship and limited opportunities push individuals to alternative paths. Furthermore, weak states often struggle to maintain control over vast territories, especially those with rough terrains, hindering their ability to quell nascent rebellions. The combination of these factors creates a breeding ground for insurgency, where the costs of rebellion are low and the potential rewards are high.

Insurgent groups create a contending sovereignty with the incumbent state [6]. While Fearon and Laitin (2001) acknowledge the role of grievances in initiating violence they argue that these factors are secondary to conditions favoring insurgency. Grievances, whether based on ethnicity, religion, or economic inequality, are widespread and do not always escalate into violence [12]. The technology of insurgency, characterized by guerrilla tactics and decentralized organization, does not necessitate widespread popular grievances or ethnic solidarity. Instead, insurgents can exploit local knowledge and weak state capacity to sustain their movement, even without broad-based support. Other scholars like Gurr (1971) argue that grievances, particularly relative deprivation (RD), are central to understanding the onset of civil war [11]. RD is the perceived discrepancy between the goods and conditions of life to which people believe they are entitled (Value Expectations) and their perceived ability to attain them (Value Capabilities). These unmet expectations can be economic, political, or social. When the gap between expectations and reality becomes too wide, it can lead to frustration, anger, and collective action against the regime. In this view, grievances act as the spark that ignites the tinder of insurgency, motivating individuals to act as revolutionary actors.

Gurr's argument builds upon the frustration-aggression hypothesis, which posits that frustration when not adequately addressed, can escalate into aggression. In the context of Violence Initation, the gap between value expectations and capabilities fuels frustration among the populace. If the state is unable or unwilling to address these grievances and bridge the gap, this frustration can boil over into aggression and collective action. Gurr refers to this as the "end of adaptation," where the state's failure to adapt and respond to the growing frustration marks the beginning of a violent process. This frustration can manifest as overt non-violent resistance, such as riots and protests, or covert bottom-up resistance. The intensity and scope of this relative deprivation determine the potential for collective violence. The more widespread and deeply felt the deprivation, the greater the likelihood of violence. If the state cannot suppress this escalating resistance, it can lead to a volatile escalation toward violent conflict.

Violence Continuation: Greed and Repression

Collier and Hoeffler (2004) provide an economic model of civil war that emphasizes the role of greed in the continuation of violence, arguing that the feasibility of funding rebellion is a key factor. Collier and Hoeffler find that the abundance of natural resources, proxied by primary commodity exports, significantly increases the risk of civil war. This suggests that the opportunity for resource exploitation can fuel and prolong conflict. The authors also find that the presence of diasporas, which can provide financial and logistical support to rebel groups, increases the risk of conflict renewal. An example is fighting for control over diamond mines in West Africa. Collier and Hoeffler's greed argument challenges the traditional grievance-based explanations of civil war, shifting the focus to the economic incentives and opportunities that sustain conflict. The authors argue that rebellions can be viewed as profit-driven enterprises, where the availability of resources and financial support determines viability. This perspective suggests that measuring economic drivers of conflict may be crucial for predicting the end of violent events.

Repression on the part of the incumbent government also determines Violence Continuation. Zhukov (2023) introduces the concept of a "threshold level of violence" [9]. This threshold represents the point at which state repression becomes so extreme that it outweighs the opposition's ability to recruit and sustain its fight. Below this threshold, state violence may backfire, fueling grievances and bolstering the rebellion. Zhukov's analysis of the conflict in Chechnya reveals that moderate levels of repression by the Russian government initially led to an increase in rebel activity [9]. However, when the government escalated its violence beyond the threshold, rebel activity sharply declined. This suggests that while repression can be a tool for the state to maintain control, its effectiveness hinges on exceeding this critical level of violence.

The threshold level of violence varies depending on several factors. Zhukov (2023) identifies government intelligence capabilities as a key determinant, an overlapping measure of opportunity [9], [13]. When the state has better information about the rebels and their supporters, it can use violence more selectively, thus lowering the threshold needed to suppress the rebellion. This is illustrated by Zhukov's finding that governments with more secret police agencies tend to have lower thresholds. Conversely, external support for rebels can raise the threshold, as it allows them to sustain their fight even under intense state repression. Other factors democratic factors of governance, such as restrictions on freedom of speech and movement, can also influence the threshold by limiting the opposition's ability to mobilize and communicate.

2 A Case Study: The French Revolution

Structural factors provide the underlying conditions that enable social revolutions to occur, even if they do not directly cause them. Revolutionary vanguards are not necessary to bring about revolution. In the cases of France, Russia, and China, the pre-revolutionary states were wealthy agrarian autocracies that faced crises when confronted by more economically developed military competitors. For France it was the Seven Years' War, for Russia it was the Russo-Japanese War and WW1, for China it was the Second Sino-Japanese War. This combination of external pressures and internal structural conditions, including conflicts of interest between landed upper classes and state rulers, contributed to the weakening and breakdown of state organizations. Peasant revolts also played key roles in all three revolutions, underscoring the role of class antagonisms between dominant and subordinate. While revolutionary vanguards drove the revolution, Skocpol argues that these vanguards did not create the crises they exploited [14]. Rather, multiple structural factors intersected to make these societies vulnerable to revolution, with the weakening of state organizations being an especially crucial element in the emergence of revolutionary situations.

In pre-revolutionary France, a combination of structural factors created a landscape ripe for revolution, although these factors alone did not directly spark the revolutionary outbreak. France's agrarian autocracy faced a severe internal fiscal crisis from the costs of participation in the Seven Years' War and the rebalancing of wealth outside the aristocracy. Taxes were primarily levied on the Third Estate commoners, exacerbating tensions between the dominant and subordinate classes [15]. Weakened state organization and made it vulnerable to revolutionary challenges from below. Venal office was the sale of office by the monarchy to rich members of the Third Estate that promised ennoblement into the Second Estate. However, these "sword nobility" at the top of the third estate became frustrated by their unacceptance in the second estate [16]. Even though they had entered the second estate through venal ennoblement, they were not respected as such by the "robe nobility" who held positions from ancestral inheritance [17]. Relative deprivation, the discrepancy between what people believe they deserve and what they can actually attain, is a key factor in the formation of contending groups that oppose the state. When people's expectations rise rapidly but the means to satisfy those expectations are lacking, frustration and anger grow. Relative deprivation of the "sword nobility" who had the economic power to contend with the First and Second Estate and became revolutionary vanguards. It was the preexisting structural landscape of class tensions, international pressures, and state weakness that made France susceptible to revolution.

Revolutions occur when contenders or coalitions advance exclusive alternative claims to control the government and gain significant commitment to those claims from the population. Tilly argues that contending groups can emerge in three ways: through the mobilization of a new contender outside the polity, a challenger turning away from accepting the polity's current rules, or an existing member of the polity turning away from its established place [6]. In France, both liberals within the First Estate, discontent "sword nobility" in the Second Estate, and poor peasants in the Third Estate emerged as challengers to the monarchy. The national assembly formed as a competing group with exclusive legislative control and an alternative vision of government that had the power to contend with the incumbent state [18], [19]. Under enabling environmental conditions such as state weakness, class antagonisms, and ideological shifts, the aggrieved group, united by their common deprivation, mobilizes into a contending group. They develop an alternative vision of governance to challenge the incumbent authorities. The emergence of this contending group, driven by their unresolved deprivations, creates a volatile potential for confrontation and political violence with the state.

3 Applying the Vision Transformer to Conflict Forecasting

The ViEWS prediction competition from Peace Research Institute Oslo and Uppsala University was an international effort in 2022 aimed at forecasting changes in state-based violence in Africa [8]. With standardized metrics across temporal, spatial, data-partitioning, and validation dimensions, the competition provides an extensive illustration of the boundaries of current research and unique insights into the architectural effectiveness of purely quantitative CNNs, LSTMs, and Markov models [2], [20], [21]. Models based on theoretically-motivated predictors such as the presence of UN peacekeeping operations performed on par or sometimes better than purely quantitative models. We use the data, evaluation, and prediction practices in the ViEWS competition to help benchmark our model against the previous literature. Furthermore, we address explainability limitations presented in the ViEWS competition by predicting a distribution of fatalities rather than a single value prediction using a Monte Carlo simulation [3]. Our model is a Transformer-attention based model with an architecture based on the Vision-Transformer (ViT). We believe that spatial-temporal encodings in the ViT model, the data augmentations to generate a larger dataset, the feature engineering to encode theory-motivated data within each token, and extending context length to outside a 12-month prediction window.

Increasing availability of high-quality and high-resolution data about political violence, and decreasing costs of compute have spurred researchers to adopt advanced neural networks from the deep learning literature to predict conflict patterns. Work from early 2022 presented approaches using Convolutional Neural Networks (CNN), Graph Neural Networks (GNN), and Long-Short Term Memory (LSTM) models [2], [21]. Each of these models excel at a specific task and are used to encode and propagate different dynamics of spatial and temporal information. They have a comparative advantage against the existing ML tradition of Random Forests, Logit, and Lasso with a higher capacity to capture the dynamics of conflict and are especially effective at predicting changes in fatalities when conflict is ongoing [3], [22]. However, they notably struggle to predict the eruption of violence in peaceful locations. Other models have incorporated strategies from the NLP tradition such as Latent Dirichlet Allocation (LDA) to identify signals of rising tensions to overcome the conflict trap [23].

The Transformer and attention are the architecture and mechanism on which all modern large language models are built. Unlike RNNs and LSTMs, the Transformer can selectively associate with any single moment within a time series, rather than having all previous states rolled up into one residual. This allows a model to better learn ephemeral temporal dynamics, which we believe will result in advanced forecasting ability as theoretical predictors of violence may not be latent sequentially. We propose a novel Transformer architecture with three strategies. 1. We use the PRIO-GRID as our basis to encode values of Relative Deprivation, State Weakness, and Existing Resistance. All values are concatenated into a 1D vector that becomes the "token" of each PRIO-GRID. 2. We adopt the positional encoding technique of the Vision Transformer (ViT) to integrate spatial dynamics [24]. A 12-month deep set of 5x5 PRIO-GRIDs are inputs into the Transformer, and positional encodings are added into PRIO-GRID tokens. Therefore, each PRIO-GRID encodes information regarding its relative position to another, but not its absolute position on a map. 3. We propose a new implementation of an Attention Sink to extend the forecasting history of our models even with a limited context length [25]. Specifically, we create special "sink" tokens that encode an average of all past data vectors encoded in each PRIO-GRID. These "sink" tokens allow the model to extend beyond our target forecasting window.

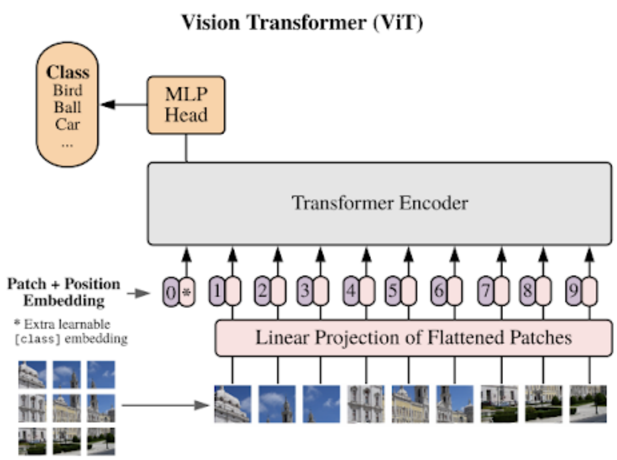

Figure 2. PRIO-GRIDs can be interpreted as images in a picture. Positional encodings are applied to spatially related different PRIO-GRIDs, although there is no absolute encoding tying PRIO-GRIDs to GPS coordinates.

We adopt and extend the datasets used in the original ViEWS 2022 forecasting competition. We use a training period from January 1990 to December 2013 and a validation period from January 2017 to December 2019. Our sources include conflict data from ACLED, UCDP, and SCAD; political indicators from Variations of Democracy and REIGN; ethnic grievances from Ethnic Power Relations (EPR); state data from World Bank; and geospatial classifications from PRIO-GRID. The conflict data distribution is right-skewed and sparse.

At the sub-national "PRIO-GRID-month" level, 99

We tie our model to our theory of revolution with our data. In each PRIO-GRID token, we encode fatalities and actors involved from the ACLED dataset. Quantitative studies have shown that previous incidents of conflict are a strong predictor of future conflict [20]. To represent Relative Deprivation, we use the Gini index, income averages of the top and bottom 20

We train on 12-month chunks with a forecasting target of month 13. We are predicting the number of casualties within a single PRIO-GRID. Since our training period only contains 156 discrete 12-month training periods, we require data-augmentation techniques to extend our data [3]. Because we choose to model our training using 5x5 PRIO-GRIDs in 12-month increments, we can extend our dataset to over 2 million separate samples due to the 15,000 discrete PRIO-GRIDs in Africa [28]. Through Gaussian noise injection, we add some variation in our 1D PRIO-GRID tensor such that our model generalizes better to small variations between countries in time. Furthermore, gaussian interpolation between data points from month to month and reorienting our PRIO-GRIDs by their positional encoding can extend our dataset to some 10-20 million samples [29].

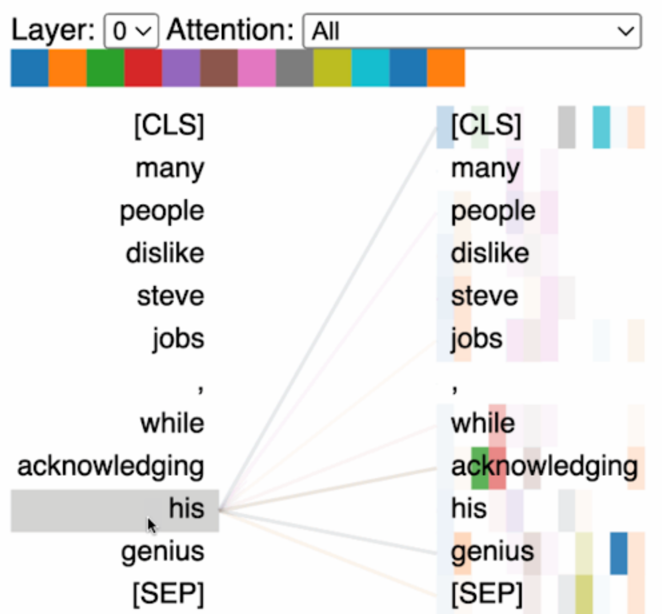

Figure 3. An attention map that maps how each token selectively attends to other tokens, and which attention head attends [30].

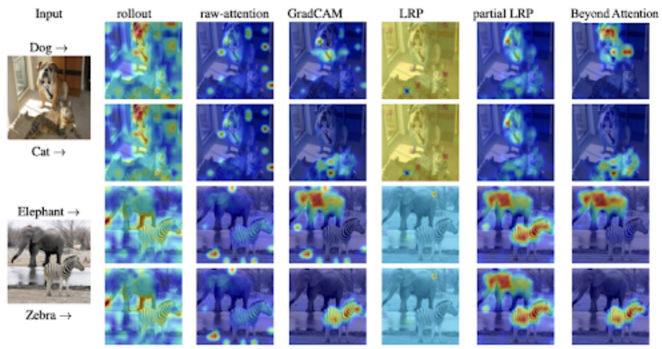

Figure 4. Various attention heat maps to explain Vision Transformers. Individual image chunks can be thought of as PRIO-GRIDs [31].

Outputting a Distribution of Fatalities

The distribution of fatalities is right-skewed and zero-inflated; When there are fatalities they are more likely near zero and sub-zero fatalities are impossible [32]. Furthermore, most PRIO-GRIDs do not have deaths encoded at the month level, leaving us with many data points that do not contribute to training. Therefore, a point-based forecast of a mean value of fatalities like in ViEWS 2022 is not sufficiently informative to balance risks between zero-valued events and high-valued events [4], [32]. We instead build a probabilistic forecast of outputs using a Monte Carlo Simulation.

We have both continuous and discrete inputs to our model, which we handle separately. For continuous inputs, we assign a normal, logarithmic, uniform, or triangular prior based on data visualization modeling. For discrete inputs, we define a Probability Mass Function (PMF) to assign probabilities for each discrete value based on our historical data [33]. We use these distributions to perform our Monte Carlo Simulation, where we randomly sample values for each input, These values are run through the model and an output distribution is created after many simulations [34].

Interpreting the Outputs of a Black Box Model

The explainability literature in deep learning is a growing field. However, the inherent complexity of scale within these models makes it difficult to model the impact of each input variable. We propose using attention maps to visualize the spatial-temporal interactions of 1D data tensors [35]. Through this method, we can visualize which tokens in the 5x5 grid in the past 12 months were the most impactful in predicting the results of the 13th month. In the past, ViTs have been bootstrapped to SHAP values such that we can determine the predictive power of any single input variable in the final output [36]. A combination of both spatial-temporal data in the 5x5x12 attention map and SHAP values to explain the 1D tensor provides both a visual and quantitative analysis to explain the forecasting results of the model. To prove our mechanism, we seek to visualize the shift from strong SHAP values in each phase of revolution over time. For example, we should see strong attention between PRIO-GRIDs of predicted violence and previous resistance. We should also see periods where PRIO-GRIDs predicting resistance have high attention patterns to previous periods where Relative Deprivation data values show high SHAP values.

4 Conclusion

We propose a novel approach to conflict forecasting by integrating a unified theory of revolution with a state-of-the-art transformer model. By drawing from the works of Skocpol, Tilly, and Gurr, we articulate a revolutionary mechanism driven by greed, grievance, opportunity, and repression, and We then demonstrate how a Vision Transformer (ViT) architecture can be used to model and forecast this mechanism by encoding its key components into input data tensors. Through data augmentation, feature engineering, and extended context length, we aim to improve predictive performance over existing benchmarks. Moreover, by leveraging attention maps and SHAP values, our model offers explainability, allowing us to visualize the shift in predictive power of different factors over the phases of violence. This approach not only advances the field of conflict forecasting but also showcases the potential of integrating social science theory with cutting-edge AI techniques. As such, it opens up new avenues for understanding and anticipating complex social phenomena.

References

[1] "ViEWS: A political violence early-warning system." Accessed: Apr. 15, 2024. [Online]. Available: https://journals.sagepub.com/doi/epub/10.1177/0022343319823860

[2] "Conflict forecasting with event data and spatio-temporal graph convolutional networks." Accessed: Apr. 15, 2024. [Online]. Available: https://www.tandfonline.com/doi/epdf/10.108

[3] H. Hegre, P. Vesco, and M. Colaresi, "Lessons from an escalation prediction competition," Int. Interact., vol. 48, no. 4, pp. 521–554, Jul. 2022, doi: 10.1080/03050629.2022.2070745.

[4] P. Vesco et al., "United they stand: Findings from an escalation prediction competition," Int. Interact., vol. 48, no. 4, pp. 860–896, Jul. 2022, doi: 10.1080/03050629.2022.2029856.

[5] X. Li et al., "Interpretable Deep Learning: Interpretation, Interpretability, Trustworthiness, and Beyond." arXiv, Jul. 15, 2022. doi: 10.48550/arXiv.2103.10689.

[6] C. Tilly, "Does Modernization Breed Revolution?," Comp. Polit., vol. 5, no. 3, pp. 425–447, 1973, doi: 10.2307/421272.

[7] J. C. Scott, Weapons of the Weak: Everyday Forms of Peasant Resistance. Yale University Press, 1985. Accessed: May 13, 2024. [Online]. Available: https://www.jstor.org/stable/j.ctt1nq836

[8] H. Hegre et al., "ViEWS: A political violence early-warning system," J. Peace Res., vol. 56, no. 2, pp. 155–174, Mar. 2019, doi: 10.1177/0022343319823860.

[9] Y. M. Zhukov, "Repression Works (Just Not in Moderation)," Comp. Polit. Stud., vol. 56, no. 11, pp. 1663–1694, Sep. 2023, doi: 10.1177/00104140231152778.

[10] P. Collier and A. Hoeffler, "Greed and Grievance in Civil War," Oxf. Econ. Pap., vol. 56, no. 4, pp. 563–595, 2004.

[11] T. R. Gurr, Why men rebel, 4. printing. in Center of international studies publications. Princeton, New Jersey: Princeton University Press, 1974.

[12] J. D. Fearon and D. D. Laitin, "Ethnicity, Insurgency, and Civil War," Am. Polit. Sci. Rev., vol. 97, no. 1, pp. 75–90, 2003.

[13] L.-E. Cederman and M. Vogt, "Dynamics and Logics of Civil War," J. Confl. Resolut., vol. 61, no. 9, pp. 1992–2016, 2017.

[14] T. Skocpol, States and Social Revolutions: A Comparative Analysis of France, Russia and China. Cambridge: Cambridge University Press, 1979. doi: 10.1017/CBO9780511815805.

[15] "French Revolution — History, Summary, Timeline, Causes, Facts — Britannica." Accessed: May 13, 2024. [Online]. Available: https://www.britannica.com/event/French-Revolution

[16] W. Doyle, Venality: The Sale of Offices in Eighteenth-Century France. Oxford University Press, 1996. doi: 10.1093/acprof:oso/9780198205364.001.0001.

[17] E. Haddad, "Nobility of the Sword, Nobility of the Robe: Social Spaces and Ideological Borders," L'Atelier Cent. Rech. Hist. Rev. Électronique CRH, no. 22 Bis, Art. no. 22 Bis, Jan. 2021, doi: 10.4000/acrh.12045.

[18] "National Assembly — Revolutionary, Estates-General, 1789 — Britannica." Accessed: May 13, 2024. [Online]. Available: https://www.britannica.com/topic/National-Assembly-historical-French-parliament

[19] T. Tackett, "Nobles and Third Estate in the Revolutionary Dynamic of the National Assembly, 1789–1790," Am. Hist. Rev., vol. 94, no. 2, pp. 271–301, Apr. 1989, doi: 10.1086/ahr/94.2.271.

[20] D. Randahl and J. Vegelius, "Predicting escalating and de-escalating violence in Africa using Markov models," Int. Interact., vol. 48, no. 4, pp. 597–613, Jul. 2022, doi: 10.1080/03050629.2022.2049772.

[21] B. J. Radford, "High resolution conflict forecasting with spatial convolutions and long short-term memory," Int. Interact., vol. 48, no. 4, pp. 739–758, Jul. 2022, doi: 10.1080/03050629.2022.2031182.

[22] R. Blair, C. Blattman, and A. Hartman, “Predicting Local Violence: Evidence from a Panel Survey in Liberia.” Rochester, NY, Oct. 20, 2016. doi: 10.2139/ssrn.2856479.

[23] H. Mueller and C. Rauh, “The Hard Problem of Prediction for Conflict Prevention,” J. Eur. Econ. Assoc., vol. 20, no. 6, pp. 2440–2467, Dec. 2022, doi: 10.1093/jeea/jvac025.

[24] A. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.” arXiv, Jun. 03, 2021. doi: 10.48550/arXiv.2010.11929.

[25] G. Xiao, Y. Tian, B. Chen, S. Han, and M. Lewis, “Efficient Streaming Language Models with Attention Sinks.” arXiv, Apr. 06, 2024. doi: 10.48550/arXiv.2309.17453.

[26] “The V-Dem Dataset – V-Dem.” Accessed: May 13, 2024. [Online]. Available: https://v-dem.net/data/the-v-dem-dataset/

[27] “World Bank Open Data,” World Bank Open Data. Accessed: May 13, 2024. [Online]. Available: https://data.worldbank.org

[28] “PRIO-GRID.” Accessed: May 13, 2024. [Online]. Available: http://grid.prio.org

[29] “Interpolation methods for time series data – d3VIEW.” Accessed: May 13, 2024. [Online]. Available: https://www.d3view.com/interpolation-methods-for-time-series-data/

[30] A. Morgan, “Explainable AI: Visualizing Attention in Transformers,” Comet. Accessed: May 13, 2024. [Online]. Available: https://www.comet.com/site/blog/explainable-ai-for-transformers/

[31] R. Kashefi, L. Barekatain, M. Sabokrou, and F. Aghaeipoor, “Explainability of Vision Transformers: A Comprehensive Review and New Perspectives.” arXiv, Nov. 12, 2023. Accessed: May 13, 2024. [Online]. Available: http://arxiv.org/abs/2311.06786

[32] H. Hegre, P. Vesco, M. Colaresi, and J. Vestby, “The 2023/24 VIEWS Prediction competition Predicting the number of fatalities in armed conflict, with uncertainty,” 2023.

[33] “Probability Mass Function — PMF.” Accessed: Jun. 03, 2024. [Online]. Available: https://probabilitycourse.com/chapter3/$3_{13p}mf.php$

[34] “Monte Carlo Simulation: What It Is, History, How It Works, and 4 Key Steps,” Investopedia. Accessed: Jun. 03, 2024. [Online]. Available: https://www.investopedia.com/terms/m/montecarlosimulation.asp

[35] J. Vig, “Visualizing Attention in Transformer-Based Language Representation Models.” arXiv, Apr. 11, 2019. doi: 10.48550/arXiv.1904.02679.

[36] I. Covert, C. Kim, and S.-I. Lee, “Learning to Estimate Shapley Values with Vision Transformers.” arXiv, Mar. 01, 2023. doi: 10.48550/arXiv.2206.05282.