Digital Humanities and Security Studies: When predictive analytics both create and prevent harm, how should its development and application be guided?

The rapid adoption of predictive analytics urgently invites critique from the Digital Humanities (DH). As Jon Chun and Katherine Elkins argue “we need humanists now more than ever”, raising an essential question, “will we program humanity into our tools, or will we cede our humanity to AI?”.[1] While critical AI studies have emerged within DH, a significant gap remains as domain-specific digital humanists are needed to evaluate socio-technical systems in many fields.[2] Nowhere is this more pertinent than in the humanitarian sector, where no international or human rights law currently govern AI use, and “half of AI researchers now believe there is a 10 per cent chance humanity will go extinct from our inability to control AI.[3]

As intergovernmental and human rights organisations increasingly turn to predictive modelling, DH offers a critical lens to assess both the design and implications of such systems. While security studies have been housed within international relations, concentrating on diplomacy, conflict resolution and human rights; this research asserts DH’s role in this discipline, addressing the complex and intricate demands of the humanitarian sector. Drawing on existing DH frameworks, such as Mark Coeckelbergh’s ‘What is digital humanism?’, this research emphasises the importance of human-centred design, alignment with human values and interdisciplinary approaches.[4] Unlike legal and policy perspectives, which primarily focus on compliance and governance, DH has the capacity to influence the technical foundations of AI security technologies. As legal scholars assert, “we international lawyers have many talents, but digital literacy is hardly one of them.”[5]

DH provides both technical methods and critical analysis to assess socio-technical systems, positioning itself as a unique interdisciplinary perspective. Furthermore, DH takes on an essential role in resolving the “expertise gap’ identified by international organisations like the UN, as the tech sector continues to interject peacebuilding efforts.[6]

---

[1] Jon Chun and Katherine Elkins, “The Crisis of Artificial Intelligence: A New Digital Humanities Curriculum for Human-Centred AI,” *International Journal of Humanities and Arts Computing* 17, no. 2 (2023), p. 149.

[2] Ibid., 149.

[3] Ibid., 161.

[4] Mark Coeckelbergh, “What is Digital Humanism? A Conceptual Analysis and an Argument for a More Critical and Political Digital (Post)Humanism,” *Journal of Responsible Technology* 17, (2024), p. 3

[5] Tommaso Soave, “From the Rule of Law to the Rule of Code? Predictive Data Analytics and the Changing Contours of UN Humanitarian Response.” Central European University (CEU), May 4, 2021, p.1

[6] Eduardo Albrecht, “Predictive Technologies in Conflict Prevention: Practical and Policy Considerations for the Multilateral System,” *UNU-CPR Discussion Paper Series*, June 2023, p. 8.

Building on this foundation, this essay critically examines the epistemic assumptions, biases, and ethical implications underpinning the design of conflict prediction models. Rather than evaluating model outputs in isolation, it interrogates the collection, categorisation, and structuring of data that shape predictive capabilities. By analysing one of the most transparent and conscientious quantitative conflict datasets, alongside the qualitative reports of global NGOs, this study highlights how different data sources construct narratives of conflict. In doing so, this essay argues that models are not neutral tools but manifestations of human decision-making, shaped by geopolitical, methodological, and ethical considerations.

Dataset Analysis

As security studies researcher, Kauffman outlines:

> "*The production and selection of data is as much based on specific decisions about what and how to measure as is the ensuing analysis that is predicated upon the questions that one wishes to address to a data set*"[7]

This observation highlights the inherent limitations of datasets built for predictive modelling; the data itself is shaped by assumptions about what patterns *should* be detected and, ultimately, what outcomes *should* be prevented. Digital humanists reinforce this issue, arguing that "data, when selected for a use-case already in mind, is always constructed, never raw or objective".[8] In the context of conflict prediction, this elicits questions about the pattern's creators are expecting to find.

For instance, data of a particular region may be captured in greater detail due to the *near-repeat hypothesis*, which speculates that events will reoccur in the same location instead of elsewhere.[9] While this reasoning is grounded in evidence-based trends, it is not unfailing. One consequence of this assumption is the overrepresentation of certain conflict-prone regions and the obfuscation of emerging threats in underreported regions. The 2011 Arab Spring provides a striking example where initial focus was on the protests in Tunisia and Egypt which resulted in delayed recognition of the events unfolding in Libya and Syria.[10] Such moments depict the constraints of data in shaping our collective understanding of conflict.
Seeking to address this, *Armed Conflict Location & Event Data (ACLED)*, attempts a more comprehensive approach. ACLED gathers global information from traditional media, international organisations & NGOs, local partners and social media (including Twitter, Telegram and Whatsapp). It then codes events based on standardised categories such as

---

[7] Kaufmann, Mareile, Simon Egbert, and Matthias Leese. "Predictive Policing and the Politics of Patterns." *The British Journal of Criminology* 59, no. 3 (May 2019), p. 678.

[8] Sarah Ciston, Zach Mann, Mark C. Marino, and Jeremy Douglass, "Can Open-Source Fix Predictive Policing? Anti-Racist Critical Code Studies Approach to Contemporary AI Policing Software," *Digital Humanities Quarterly* 19, no. 1 (2025), p. 6.

[9] Ibid., 5.

[10] Ekaterina Stepanova, "The Role of Information Communication Technologies in the 'Arab Spring': Implications Beyond the Region," *PONARS Eurasia Policy Memo* No. 159, May 2011, Institute of World Economy and International Relations.

‘event type’, ‘actors involved’, ‘location’, and ‘time.’
Yet, this method is not without flaws. In regions with low-press freedom, conflicts may be underreported or misrepresented and in active conflict zones the only available source may be a corrupt government or an armed group itself – which inevitably introduces risks of strategic misinformation.
Such vulnerabilities are even more pronounced in alternative open-data initiatives such as the Global Database of Events, Language and Tone (GDLET) which aggregates global news media in real time. While GDLET can provide imminent access to unfolding events (updating every fifteen minutes), unlike ACLED, it does not verify reports before categorising and geolocating them. This introduces significant risks, for instance, reliance on GDLET for conflict prediction in the Gaza-Israel context can easily become deeply problematic, given Israeli government’s use of bot farms to spread disinformation.[11]
However, even when datasets employ verification mechanisms, like ACLED, conflict-induced blackouts and poor telecommunications infrastructure, as seen in Gaza can still skew results. As even the best methods of collection cannot remove digital global disparities. Technology policy researchers note that inequal structures and histories tend to result in inequal outputs:

> *“Inequality hangs over the past like fog. It lingers for generations. It affects every facet of people’s lives, and it seeps into the data about them*.”[12]

This observation encapsulates a fundamental issue with conflict prediction – data is not neutral. It reflects and perpetuates inequalities, meaning that the ‘gaps’ in data often reinforce very real gaps in lived experience.

Beyond data collection, further complications arise in data preprocessing. Standardised categories are employed to organise data and identify trends and yet the decision-making for this is often hidden by the end product. For example, when determining the actors involved in a conflict, who decides when groups or individuals are deemed ‘rebels’ vs. ‘freedom fighters’? What happens when a model is trained to make that choice? Ultimately such decisions become embedded into the model’s architecture, concealing biases to practitioners who need to act on the predictions.[13]
Furthermore, other data-providers are not necessarily as forthcoming with their methods as ACLED and GDLET are. For example, *The Violence & Impacts Early-Warning System (ViEWS)* does not share their collection processes or methodology. This is of even greater concern as unlike ACLED and GDLET, ViEWS provides free access to its predictive model, forecasting conflict up to three years ahead. As ViEWS was initially developed to monitor and predict political violence across Africa, there is concern that its probabilistic assessments are based on an abundance of data from historically volatile regions. This introduces the

---

[11] Nick Robins-Early, “OpenAI Says Russian and Israeli Groups Used Its Tools to Spread Disinformation,” *The Guardian*, May 30, 2024.
Sheera Frenkel, “Israel Secretly Targets U.S. Lawmakers with Influence Campaign on Gaza War,” *The New York Times*, June 5, 2024.

[12] Hideyuki Matsumi and Daniel J. Solove, “The Prediction Society: AI and the Problems of Forecasting the Future,” *Legal Studies Research Paper* No. 2023-58, *GWU Law School Public Law Research Paper* No. 2023-58 (2025), p. 25.

[13] Kaufmann, Egbert, and Leese, “Predictive Policing and the Politics of Patterns,” 678.

‘Fossilisation Problem’, where algorithmic predictions make it harder for people to escape from the past.[14] By presuming that ‘the past repeats itself and thus the future will be similar to the past’, algorithmic forecasts can reinforce existing assumptions as regions constantly labelled ‘high risk’ might be treated as if conflict is inevitable - even if the evidence of escalation is weak.[15] For example, low-stakes indicators such as minor protests might be interpreted as signs of impending conflict, meanwhile a state that has regular and consistent high-stakes protests and state-based violence that do not materialise in war; will consequently be depicted with low-conflict risk.

While it is not possible to discern which countries ViEWS’ present forecast may be ‘fossilising’, we can look to examples of the past to better understand how this can transpire. After the Good Friday Agreement, Northern Ireland saw a decline in violence, falling rates of unemployment and GDP growth by 25%; outcomes that would not have been predicted in the early 2000s.[16] Similarly, the post-Suharto years of Indonesia have seen democratic stability and Rwanda has become one of the most stable nations in East Africa, marking 30 years since its Genocide in 2024. In each of these cases, past instability did not indicate future conflict, challenging the core principles underpinning predictive models.

Moreover, these models may fail to identify emerging risks in seemingly stable states too. This is best illustrated by ViEWS forecast of the US. Predicting the states to be in the lowest category of conflict-risk:

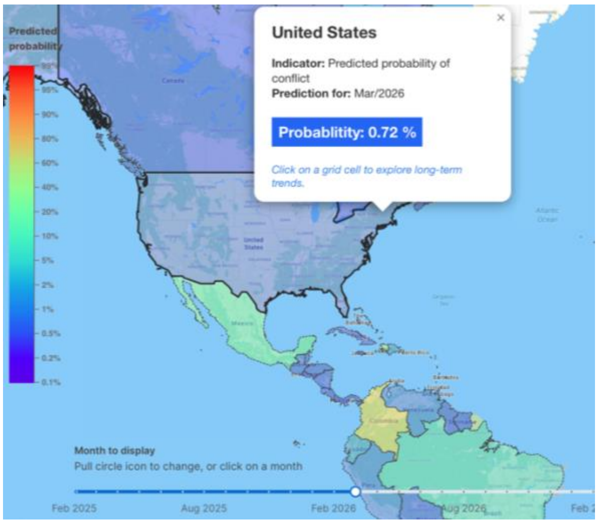

Figure 1

This prediction contradicts recent trends, given the escalations of domestic political insurrections and violence occurring in the US over the last few years. Most notably, the self-coup attempted by Trump supporters at Capitol in 2021, demonstrating willingness of militia groups to use force to challenge democratic institutions. In addition to this, the US has seen a

---

[14] Matsumi and Solove, “The Prediction Society,” p. 22.

[15] Ibid., p. 21.

[16] Graham Brownlow, David Jordan, and John Turner, “The Good Friday Agreement at 25: Has There Been a Peace Dividend?” *Economic Observatory*, April 3, 2023

rise in violent threats against officials – including Paul Pelosi in 2022 and the presidential assassination attempt in July 2024. Furthermore, the Department of Homeland Security has classified the US at 'high-risk' of domestic terrorism, citing the "domestic sociopolitical developments".[17] ViEWS' failure to capture this context suggests that its risk assessments are shaped by historical *expectations* rather than realities.

Such examples expose the fundamental limitations of predictive modelling and the complexities of conflict that cannot always be distilled into quantifiable patterns. This raises critical questions surrounding predictive models. For instance, can declarations of peace (e.g. the Good Friday Agreement) be incorporated into models? If not, how do we use models alongside events that cannot be quantified to anticipate future risk?
More broadly, all the databases examined here reflect greater challenges in conflict forecasting. How is global inequality accounted for in data collection practices? How can disinformation be navigated? The UNs review of predictive technologies also highlighted that all of these databases are led by the Global North; emphasising the need for a "centralised repository of different data technologies for peacebuilding in the digital era".[18] These concerns undermine the notion that patterns are inherently factual. For this reason, legal researchers, Matsumi and Solove describe predictive models as "a dangerous and irresponsible way to create the future". [19] Their warning adjudge such tools as a means for suppression, stressing that "acting on predictions to shape the future is a profound exercise of power, and the potency of this power is often not fully appreciated."[20] Thus, not only should the production of data, design of models and wider structures be interrogated, but rather, the overall practice of conflict prediction.

Case Study

Data Collection and Preprocessing

> The primary quantitative dataset used in this study, is the ACLED Middle East dataset, which provides georeferenced data on political violence and protests. ACLED was selected due to its greater transparency in attempting to mitigate bias in comparison to other data providers outlined above. The selection of the Middle East dataset was motivated by ACLED's 2024 conflict index, which identified three out of the ten most-at-risk states as Middle Eastern (Palestine, Syria and Lebanon). To assess the predictive validity and justification of such models, this study trained a machine learning model on ACLED data from 2015 to 2022 and tested against events from 2023 onward. However, every stage of data processing introduces choices that can

---

[17] U.S. Department of Homeland Security, *DHS' 2025 Homeland Threat Assessment Indicates the Threat of Domestic and Foreign Terrorism in the Homeland Remains High*, October 2, 2024

[18] Eduardo Albrecht, "Predictive Technologies in Conflict Prevention: Practical and Policy Considerations for the Multilateral System." *UNU-CPR Discussion Paper Series*, June 2023, p. 12.

[19] Matsumi and Solove, "The Prediction Society," p. 33.
[20] Ibid., p. 33.

fundamentally shape the model's outputs, in subtle but significant ways. This study categorised events into three severity levels:

- High Severity (2): Battles, Explosions/Remote violence, Violence against civilians.
- Medium Severity (1): Riots.
- Low Severity (0): Protests, Strategic developments.

The model was trained using ACLED's 'event type' column as the primary variable but an alternative approach could have prioritised the 'disorder type'. This would have grouped 'protests' and 'riots' together, creating different patterns of classifications and interpretation. Such decisions, while appearing technical, have profound implications, altering how conflict is framed and whose actions are emphasised or downplayed. As the UN notes "it is a mistake to assume that all stakeholders share the same definitions, categories, and parameters which are fundamental to how the technology is calibrated."[21]
By outlining this process, this study does not claim objectivity, but instead highlights its own biases, asserting that transparency does not eliminate embedded assumptions. Even beyond the decisions made in this study and those made by ACLED, the reliance on third-party news sources inevitably shapes how events are defined and recorded. This reinforces the over-arching argument of this essay, that even at the earliest stage of preprocessing, data design is neither neutral nor impartial. Each decision, regardless of how well-reasoned reflects and reproduces pre-existing structures and beliefs.

Furthermore, it is worth considering what 'event types' are absent. Indicators such as 'unusual troop movements', 'increased weapon production' and the 'imposition of civilian restrictions' are not recognised conflict precursors. Alongside this, other key indicators of conflict include: illegal mining surges, government arming of civilian groups, dramatic shifts in nationalist symbols and crackdown on dual-citizens and foreign residents - all of which are not included in this dataset, while simultaneously having occurred in Israel.[22] [23] [24] [25]As legal scholarship stresses, "the data that is not available is often the product of deliberate choices".[26]

To further uncover the implications of this, this study explores qualitative sources such as the 2015 - 2025 annual reports from Human Rights Watch (HRW) and

---

[21] Albrecht, "Predictive Technologies in Conflict Prevention", p. 10.

[22] Heidelberg Cement, *The Israeli Exploitation of Palestinian Natural Resources: Part II, Who Profits*, November 2016

[23] Scott Neuman and Eleanor Beardsley, "Israel Is Trying to Arm More Citizens with Guns Since the Hamas Attack," *NPR*, December 6, 2023

[24] Jamal Kanj, "Israeli Invention of National Symbols," *Al Mayadeen English*, June 25, 2023

[25] Yolande Knell, "Israel Passes Law to Revoke Israeli Arab Attackers' Citizenship," *BBC News*, February 15, 2023

[26] Matsumi and Solove, "The Prediction Society," pp. 25-26.

Amnesty International, contrasting the shortfall of nuance in ACLED's quantitative dataset. As organisations such as the UN express the need for contextual factors to accurately predict trends, these reports can depict greater intricacy in conflict escalation.[27] This is not to suggest they are without limitations. For instance, who decides which states have greater coverage or analysis? Are grassroots organisations involved in sharing local narratives? To what extent do HRW and Amnesty International reinforce Western assumptions? As criminologists assert:

> *"All datasets are necessarily limited representations of the world that must be imagined as such to produce the meaning they purport to show"*[28]

Thus, even with greater context, the use of qualitative data should not be mistaken as an infallible approach to conflict modelling. Furthermore, to adequately prepare the data, qualitative sources must undergo preprocessing to fit the model's computational nature. For example, the HRW and Amnesty reports were:

- Extracted and cleaned from PDFs using the PyPDF2 library – this included removal of punctuation.
- Application of Natural Language Processing (NLP) techniques:
    - Tokenisation and Lemmatisation to capture linguistic variations,
    - Named Entity Recognition (NER): Identifying geopolitical entities (GPE)

Narrative accounts are reduced to numerical correlations and keyword frequencies, detached from their broader socio-political context. This occurs without even critically interrogating the role of Amnesty and HRW in shaping conflict narratives or questioning whether these sources, despite their reputations, offer a comprehensive representation of events.

2. AI Model Development and Implementation

2.1 Feature Engineering for ACLED Data

Key features were engineered to improve predictive capabilities, for instance temporal features were used to capture the month, day of the week, and quarter of the year to capture seasonal and cyclical patterns. Similarly, geospatial and actor-based features were used for the following:

- Location-Based Conflict Intensity: Historical conflict intensity per location.
- Severity Averaging: Mean severity score for past events at each location.
- Actor Involvement: Frequency and severity of past events involving specific actors.
- Country-Level Aggregates: Summarising event frequency and severity at a national level.

[27] Ibid., p. 7.
[28] Kaufmann, Egbert, and Leese, "Predictive Policing and the Politics of Patterns," p. 678.

As the analysis of preprocessing decisions unfolds, it becomes increasingly evident that the model is being designed to recognise only the patterns deemed significant. Consequently, its ability to predict conflict is constrained by the parameters established. Events that fall outside of these predefined classifications, will not be detected. As Digital Humanists note “systems which rely on patterns only capture offenses that follow the rules upon which that algorithm relies”.[29] Yet, textual data, is not immune to this issue either, as it too, is structured according to varied assumptions.

2.2 Analysis of HRW and Amnesty Reports

For example, lexical analysis was employed with a predefined lexicon of conflict-related keywords to assess patterns.* This included:

- Conflict terms: e.g., war, violence, unrest.
- Political instability: e.g., coup, repression, dictatorship.
- Terrorism and insurgency: e.g., extremism, armed group, proxy war.

While this approach attempts to incorporate nuance with the use of qualitative data, it must still comply with the inherent structure of algorithmic design. What happens when a new form of conflict arises, but the model lacks the vocabulary to recognise it? What if sources cannot be translated from a language the system does not support? As Matsumi and Solove highlight, “applying a more rigorous scientific method to algorithmic predictions isn’t enough”.[30] On the surface, expanding language models or updating keyword lists might appear to be a simple remedy. However, the fundamental issue lies in the design choices that dictate whose perspectives are prioritised and whose realities are excluded. These are questions that require a ‘humanist approach’ rather than purely technical adjustments.[31]

3. Model Training and Evaluation

3.1 ACLED Model Training

Four machine learning models were tested:

- Random Forest Classifier: Ensemble-based, interpretable, handles imbalanced data. (Best performer; F1 Score: 0.9184, Accuracy: 0.9380)
- Gradient Boosting Classifier: Sequential boosting for refined decision trees.
- XGBoost Classifier: Efficient gradient boosting, robust to missing data.
- LightGBM Classifier: Faster tree-based boosting, optimised for large datasets.

The model demonstrated high precision (1.00) and recall (1.00) for low-severity conflict events, indicating it effectively identifies stable regions. Similarly, high-

[29] Ciston, Mann, Marino, and Douglass, “Can Open-Source Fix Predictive Policing?” p. 8.
[30] Matsumi and Solove, “The Prediction Society,” p. 51.
(see Appendix B for complete list).
[31] Ibid., p. 55.

severity conflict events were also predicted well (F1 score: 0.95). Yet, blind sports were unveiled with medium-severity events which had the weakest performance, (recall: 0.17), meaning many medium-severity conflicts were misclassified, primarily as high-severity events.

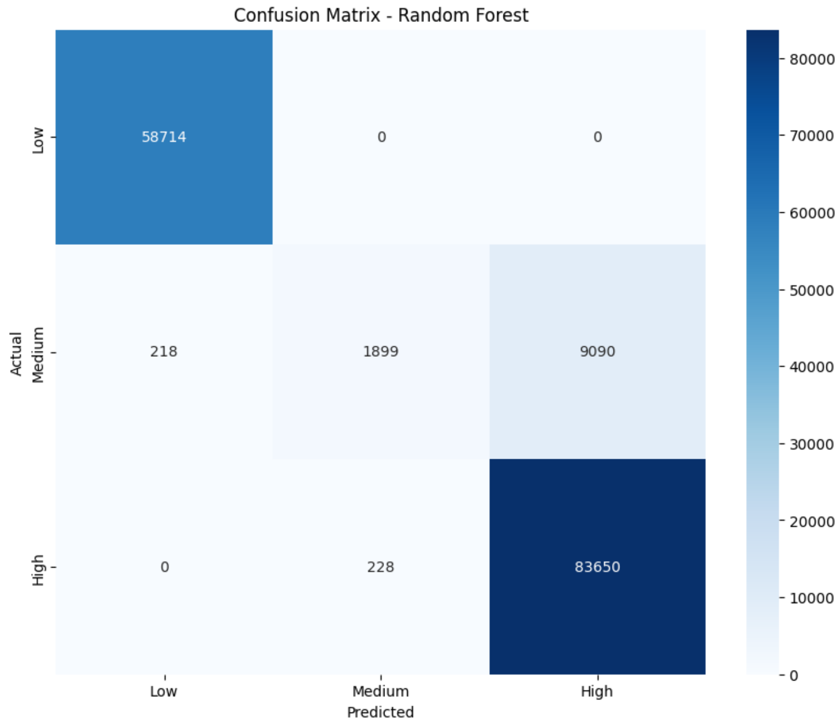

Figure 2

This reflects the difficulty in capturing political instability outside of extreme or existing cases. Kauffman, Egbert and Leese's aptly capture this issue:

> "*Patterns can only capture offenses that follow rules […] In fact, designers and programmers acknowledge (to their regret) that any behaviour that does not follow a pattern cannot be detected"*[32]

This is further illustrated in the temporal analysis as the model's accuracy varied over time, with notable fluctuations:

- January 2023 - August 2023: Accuracy ranged between 0.89 and 0.935.
- October 2023: Accuracy dropped to 0.925 following the October 7th escalation in Palestine-Israel.
- May 2024: Accuracy declined to 0.94 amid the Israeli military offensive in Rafah.
- January - February 2025: Accuracy decreased to 0.925-0.92, coinciding with the ceasefire period which commenced on the 19th January.

[32] Kaufmann, Egbert, and Leese, "Predictive Policing and the Politics of Patterns," 684.

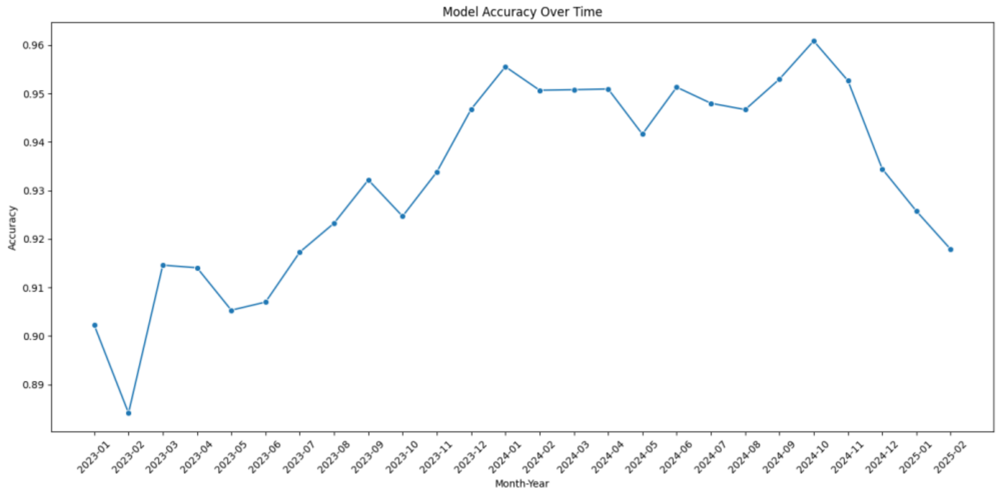

Figure 3

The geopolitical disparities also reinforce this as there was lower accuracy in Palestine (<0.8). Meanwhile low-conflict states like Kuwait, Oman, and Qatar had perfect accuracy (1.0). This suggests that prediction accuracy for conflict patterns in volatile zones – even when continuous can still be less reliable. As professor of Digital Diplomacy, Corneliu Bjola asserts:

> "*AI may not be able to completely dissolve "the fog of war", but they may be able to provide sufficient or actionable confidence in the value of the information used for taking decisions in times of crisis. To do this, an AI model need to take into consideration the factors that can blur crisis signalling and reduce the level of uncertainty that they induce as much as possible."*[33]

In other words, for an AI model to be truly effective, it must be able to process vast amounts of information including real-time data – and even then, it would only be sufficient as an assistant to human intelligence. While ACLED's extensive database includes an exhaustive list of international and local partners, its approach to responsible verification ultimately hinders it applicability for real-time intelligence. This is not to suggest, that the alternative is a more strategic solution either. As established earlier in this essay, without effective processes for source validation, models can easily become exploited by being trained on disinformation. Essentially, this issue stresses the need for diversity and quality in data, not starkly an abundance of it.

3.2 HRW and Amnesty Report Analysis

In comparison to the ACLED model, performance varied significantly between the two qualitative data-trained models. HRW had an accuracy of 50% with precision and

---

[33] Corneliu, B., "Artificial Intelligence and Diplomatic Crisis Management: Addressing the 'Fog of War' Problem." *DigDiploROx Working Paper No. 6*, University of Oxford, July 2022, p.5.

recall skewed toward conflict classification. Meanwhile, the Amnesty trained model had an accuracy of 80% that was more balanced.
This is likely due to Amnesty's concentration on systematic patterns of human rights violations, which provides a more structured documentation for the model to identify patterns. In comparison, HRW reports are narrative focused with detailed investigative journalism and selective case studies. While it is intuitive that models trained solely on qualitative data would not be sufficient at prediction, these examples can highlight the inadequacy of models trained exclusively on quantitative data.

3.3 HRW

For instance, HRW's most mentioned countries were the US, Syria, China, Russia and Ukraine, contrasting the ACLED index which labels both the US and China as 'low/inactive' and Palestine, Myanmar, Mexico, Nigeria, Brazil, Lebanon, Sudan, Cameroon and Colombia as the most extreme risks of conflict alongside Syria. While this does suggest a Western-centric approach, these results also reflect HRWs broader human rights lens which includes systemic repression and abuses within perceived stable states – often overlooked by quantitative approaches.

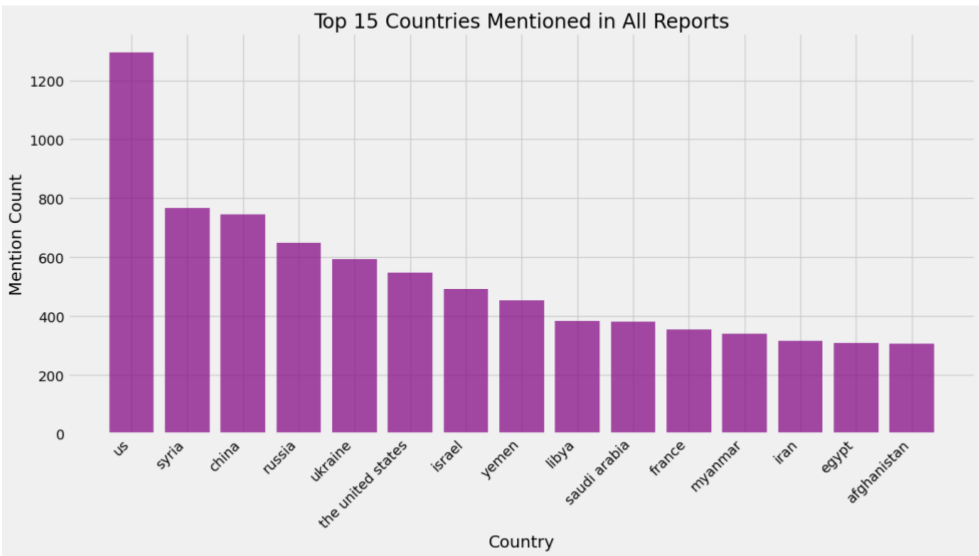

Figure 4

The most frequent keywords which included:

- violence (<4000)
- military (<3000)
- torture (2000+)
- asylum (1750)
- war (<1500).

With other significant terms including 'corruption', 'terrorism', 'disappearances', 'human rights abuses', 'repression', 'crackdown' and 'armed conflict'. While these keywords generally mirror the indicators analysed in quantitative datasets, 'repression', 'crackdown', 'corruption' and 'disappearances' correspond with notions of authoritarianism and state control.

Datasets like ACLED do not track developing signs of authoritarianism as an indicator for conflict, instead opting to record signs of state resistance. Yet, acts of state-led repression, and corruption can serve as early-warning signs of conflict. This is further reinforced by the heatmap analysis, which highlights the following keyword correlations:

    - "Torture" and "armed conflict" are strongly associated with reports from 2017 onward.
    - "State violence" has a moderate correlation with overall report length.
    - "Discrimination" shows increasing prominence over time.
- Geopolitical associations:
    - The US, Tabqa (Syria), and Khorasan Province (Afghanistan/Iran) feature in strong correlations with conflict-related keywords.

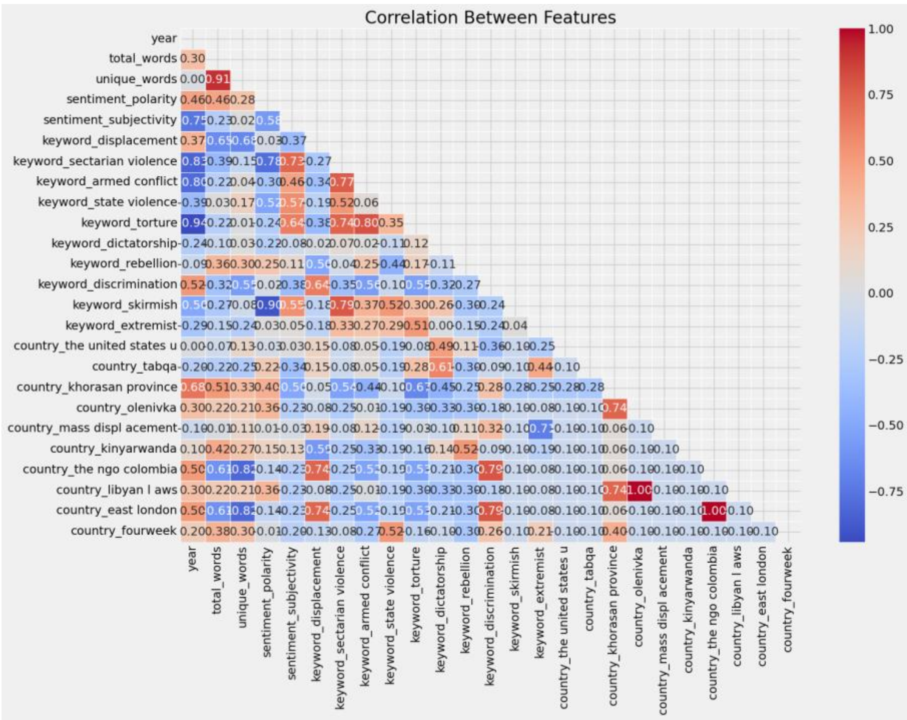

Figure 5

The consistent correlation of ‘state violence’ throughout the report emphasises its relevance in conflict, yet it is often invisible in quantitative models because it doesn’t *always* escalate into open warfare. This is a hard notion to oppose in predictive analytics as what is “not based on patterns cannot be forecasted and [therefore] has to be excluded a priori.”[34] Furthermore, authoritarian regimes are not only more likely to suppress evidence of civil resistance but may also prevent such resistance from occurring in the first place. By systematically controlling information and silencing dissent, these states can obscure signs of instability until extreme events, such as the 2011 Arab Spring uprisings, make repression visible. These paramount events can

[34] Kaufmann, Egbert, and Leese, “Predictive Policing and the Politics of Patterns,” p. 684.

then be inadequately depicted as *unforeseen* by quantitative models. Therefore, demonstrating the significance of the indicators that emerge from qualitative sources.

3.4 Amnesty

Amnesty's most mentioned countries were Syria (675), Ukraine (600), Israel (550), Russia (500) and Libya (480). In comparison to HRW, China had 350 mentions, and the US was not frequently mentioned at all.

The most frequent keywords included: violence (5500), torture (4000), military (2800), discrimination (2200), conflict (1500). With other significant terms including: asylum, war, refugee, disappearances, war crimes, terrorism, repression, crackdown, displacement.

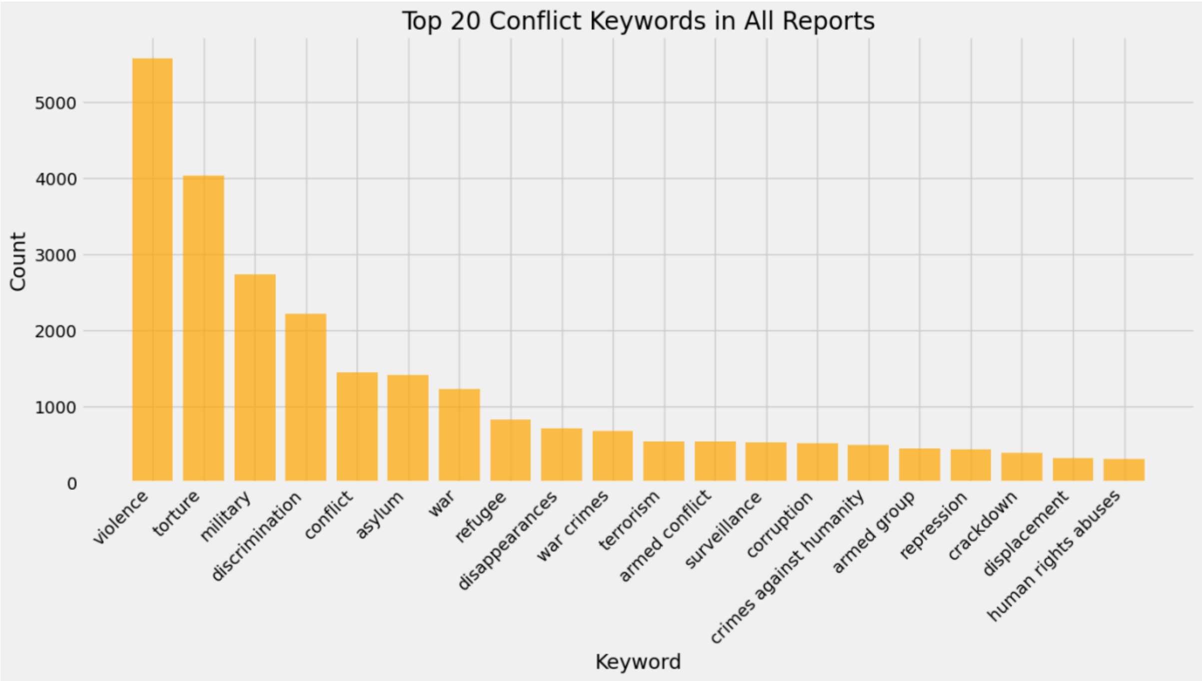

Figure 6

In comparison to HRW, the Amnesty data suggests a greater focus on active conflict zones and the humanitarian impact rather than conflict precursors. This is further reinforced in the keyword correlation:

  - "Torture" and "humanitarian crisis" highly correlated with total word count.
  - "Extrajudicial killing" moderately correlated with sentiment shifts.
  - "Rebellion" fluctuates unpredictably.
- Meanwhile, geopolitical associations
  - Mentions of North Korea, Sichuan (China) and Thailand show strong correlations with conflict-related terms.

Yet, the correlation between North Korea, Sichuan and Thailand alongside conflict keywords, suggests invisible repression and long-standing tension that models often overlook due to the focus on high-intensity violence. This reinforces the broader argument that while quantitative datasets like ACLED provide structured, event-driven conflict tracking, they often overlook the slow-building precursors of instability. In contrast, qualitative datasets capture underlying tensions and patterns of authoritarian control that *can* serve as early warning signals, while remaining difficult to quantify. Although integrating diverse data sources could benefit the scope of prediction models, these tools alone cannot replace human judgement. As research on 'automation bias' suggests, humans have a documented tendency to be less critical of suggestions made by automated decision-making systems.[35]
More fundamentally, however, we must critically assess whether predictive models can truly serve the best interests of humanity or whether their increasing integration into decision-making processes, risks legitimising abdication of human responsibility altogether.

Implications

Given this study's exploration of conflict prediction using Digital Humanist principles, it argues that if such tools must exist, they should be built on diverse, critically examined data sources. Those who rely on predictive models must have record of the decisions that have been made to label, categorise and standardise entries. Furthermore, the overall process of collection, categorisation, and structuring must include interdisciplinary perspectives.

Yet, regardless of efforts to refine and perfect these models, this essay asserts that predictive analytics should never usurp human decision-making. While this may seem intuitive within the humanitarian sector, the widespread adoption of predictive technologies in military operations underscores the urgency of this discussion. The use of AI-driven targeting systems, such as Israel's 'Lavender project' exemplifies how statistical predictions can be weaponised, withdrawing agency from human judgement. One IDF intelligence officer admitted that they had more faith in a 'statistical machine' than a grieving soldier, while another stated:

> *"I would invest 20 seconds for each target at this stage,and do dozens of them every day. I had zero added-value as a human, apart from being a stamp of approval. It saved a lot of time."*[36]

This reveals the acute dangers of normalising and thereby, legitimising computational decision-making over human sovereignty. If we accept predictive models as authoritative, we risk transferring agency in ways that may be difficult to reverse. We must critically examine how these tools influence decision-making ensuring human oversight and accountability. This essay aligns with Digital Humanist critique, advocating for transparency in conflict prediction models, including public access to the underlying code.[37] It further echoes legal

[35] Albrecht, "Predictive Technologies in Conflict Prevention", p. 8.

[36] Bethan McKernan and Harry Davies, "'The Machine Did It Coldly': Israel Used AI to Identify 37,000 Hamas Targets," *The Guardian*, April 3, 2024.

[37] Ciston, Mann, Marino, and Douglass, "Can Open-Source Fix Predictive Policing?" p. 10.

scholarship calling for an independent regulatory body, to oversee algorithmic predictions, ensuring ethical oversight and accountability.[38]

Ultimately, beyond refining these models, further discourse might explore whether analytics should be used for conflict at all. As this study has demonstrated, these tools are not neutral arbiters of risk but reflections of human judgment, inherently shaped by the assumptions and beliefs of their creators. If conflict prediction models are to be deployed, it must be with extreme caution, full transparency, and a commitment to serving - not replacing, human expertise.

[38] Matsumi and Solove, “The Prediction Society,” p. 51.

Bibliography

Primary Sources

Reports

Amnesty International. "Amnesty International Report: The State of the World's Human Rights 2014/15." London: Amnesty International, 2015. Accessed March 2025. https://www.amnesty.org/en/documents/annual-report-2014-15/.

Amnesty International. "Amnesty International Report: The State of the World's Human Rights 2015/16." London: Amnesty International, 2016. Accessed March 2025. https://www.amnesty.org/en/documents/annual-report-2015-16/.

Amnesty International. "Amnesty International Report: The State of the World's Human Rights 2016/17." London: Amnesty International, 2017. Accessed March 2025. https://www.amnesty.org/en/documents/annual-report-2016-17/.

Amnesty International. "Amnesty International Report: The State of the World's Human Rights 2017/18." London: Amnesty International, 2018. Accessed March 2025. https://www.amnesty.org/en/documents/annual-report-2017-18/.

Amnesty International. "Amnesty International Report: The State of the World's Human Rights 2019/20." London: Amnesty International, 2020. Accessed March 2025. https://www.amnesty.org/en/documents/annual-report-2019-20/.

Americas

Africa

Asia Pacific

Eastern Europe & Central Asia

Middle East and North Africa

Western Europe

Amnesty International. "Amnesty International Report: The State of the World's Human Rights 2020/21." London: Amnesty International, 2021. Accessed March 2025. https://www.amnesty.org/en/documents/annual-report-2020-21/.

Amnesty International. "Amnesty International Report: The State of the World's Human Rights 2021/22." London: Amnesty International, 2022. Accessed March 2025. https://www.amnesty.org/en/documents/annual-report-2021-22/.

Amnesty International. "Amnesty International Report: The State of the World's Human Rights 2022/23." London: Amnesty International, 2023. Accessed March 2025. https://www.amnesty.org/en/documents/annual-report-2022-23/.

Amnesty International. “Amnesty International Report: The State of the World’s Human Rights 2023/24.” London: Amnesty International, 2024. Accessed March 2025. https://www.amnesty.org/en/documents/annual-report-2023-24/.

Human Rights Watch. “World Report 2025: Events of 2024”. New York: Human Rights Watch, 2025. Accessed March 2025. https://www.hrw.org/world-report/2025

Human Rights Watch. “World Report 2024: Events of 2023”. New York: Human Rights Watch, 2024. Accessed March 2025. https://www.hrw.org/world-report/2024

Human Rights Watch. “World Report 2023: Events of 2022.” New York: Human Rights Watch, 2023. Accessed March 2025. https://www.hrw.org/world-report/2023.

Human Rights Watch. “World Report 2022: Events of 2021.” New York: Human Rights Watch, 2022. Accessed March 2025. https://www.hrw.org/world-report/2022.

Human Rights Watch. “World Report 2021: Events of 2020.” New York: Human Rights Watch, 2021. Accessed March 2025. https://www.hrw.org/world-report/2021.

Human Rights Watch. “World Report 2020: Events of 2019.” New York: Human Rights Watch, 2020. Accessed March 2025. https://www.hrw.org/world-report/2020.

Human Rights Watch. “World Report 2019: Events of 2018.” New York: Human Rights Watch, 2019. Accessed March 2025. https://www.hrw.org/world-report/2019.

Human Rights Watch. “World Report 2018: Events of 2017.” New York: Human Rights Watch, 2018. Accessed March 2025. https://www.hrw.org/world-report/2018.

Human Rights Watch. “World Report 2017: Events of 2016.” New York: Human Rights Watch, 2017. Accessed March 2025. https://www.hrw.org/world-report/2017.

Human Rights Watch. “World Report 2016: Events of 2015.” New York: Human Rights Watch, 2016. Accessed March 2025. https://www.hrw.org/world-report/2016.

Human Rights Watch. “World Report 2015: Events of 2014.” New York: Human Rights Watch, 2015. Accessed March 2025. https://www.hrw.org/world-report/2015.

U.S. Department of Homeland Security. *DHS’ 2025 Homeland Threat Assessment Indicates the Threat of Domestic and Foreign Terrorism in the Homeland Remains High*. October 2, 2024. Accessed March 2025. https://www.dhs.gov/archive/news/2024/10/02/dhs-2025-homeland-threat-assessment-indicates-threat-domestic-and-foreign-terrorism.

Data

Armed Conflict Location & Event Data Project (ACLED). “ACLED Conflict Index: December 2024 Update.” Last modified December 2024. Accessed March 2025. https://developer.acleddata.com/dashboard/conflict-index-tool/.

Armed Conflict Location & Event Data Project (ACLED). “ACLED: The Armed Conflict Location & Event Data Project [Middle East].” Accessed March 2025. https://acleddata.com/

Armed Conflict Location & Event Data Project (ACLED). *ACLED History.* ACLED, 2022. https://acleddata.com/acleddatanew/wp-content/uploads/dlm_uploads/2021/11/ACLED-History_v2_February_2022.pdf.

ViEWS (Violence Early Warning System), "ViEWS Forecast Model," screenshot, [March, 2025] https://data.viewsforecasting.org/

Figure 1

Statista Research Department. "People Shot to Death by U.S. Police 2017-2024, by Race." *Statista*, February 6, 2025. Accessed March 2025. https://www.statista.com/statistics/585152/people-shot-to-death-by-us-police-by-race/.

News Articles

Brownlow, Graham, David Jordan, and John Turner. "The Good Friday Agreement at 25: Has There Been a Peace Dividend?" *Economic Observatory*, April 3, 2023. Accessed March 2025. https://www.economicsobservatory.com/the-good-friday-agreement-at-25-has-there-been-a-peace-dividend.

Frenkel, Sheera. "Israel Secretly Targets U.S. Lawmakers with Influence Campaign on Gaza War." *The New York Times*, June 5, 2024. Archived from the original on June 8, 2024. Accessed March 2025. https://www.nytimes.com/2024/06/05/technology/israel-campaign-gaza-social-media.html.

Kanj, Jamal. "Israeli Invention of National Symbols." Al Mayadeen English, June 25, 2023. Accessed March 2025. https://english.almayadeen.net/articles/opinion/israeli-invention-of-national-symbols.

Knell, Yolande. "Israel Passes Law to Revoke Israeli Arab Attackers' Citizenship." *BBC News*, February 15, 2023. Accessed March 2025. https://www.bbc.com/news/world-middle-east-64654634.

McKernan, Bethan, and Harry Davies. "'The Machine Did It Coldly': Israel Used AI to Identify 37,000 Hamas Targets." *The Guardian*, April 3, 2024. Accessed March 2025. https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes.

Neuman, Scott, and Eleanor Beardsley. "Israel Is Trying to Arm More Citizens with Guns Since the Hamas Attack." *NPR*, December 6, 2023. Accessed March 2025. https://www.npr.org/2023/12/06/1216088371/guns-israel-hamas-gaza.

Robins-Early, Nick. "OpenAI Says Russian and Israeli Groups Used Its Tools to Spread Disinformation." *The Guardian*, May 30, 2024. Accessed March 2025. https://www.theguardian.com/technology/article/2024/may/30/openai-disinformation-russia-israel-china-iran.

Scahill, Jeremy. "Netanyahu's War on Truth: Israel's Ruthless Propaganda Campaign to Dehumanize Palestinians." *The Intercept*, February 7, 2024. Accessed March 2025.

https://web.archive.org/web/20240208172324/https://theintercept.com/2024/02/07/gaza-israel-netanyahu-propaganda-lies-palestinians/.

Secondary Sources

Journal Articles

Chun, Jon, and Katherine Elkins. "The Crisis of Artificial Intelligence: A New Digital Humanities Curriculum for Human-Centred AI." *International Journal of Humanities and Arts Computing* 17, no. 2 (2023): 147-167.

Ciston, Sarah, Zach Mann, Mark C. Marino, and Jeremy Douglass. "Can Open-Source Fix Predictive Policing? Anti-Racist Critical Code Studies Approach to Contemporary AI Policing Software." *DHQ: Digital Humanities Quarterly* 19, no. 1 (2025): 1-50.

Coeckelbergh, Mark. "What is Digital Humanism? A Conceptual Analysis and an Argument for a More Critical and Political Digital (Post)Humanism." *Journal of Responsible Technology* 17 (2024): 100073.

Stepanova, Ekaterina. "The Role of Information Communication Technologies in the 'Arab Spring': Implications Beyond the Region." *PONARS Eurasia Policy Memo* No. 159, May 2011. Institute of World Economy and International Relations (IMEMO), Russian Academy of Sciences.

Kaufmann, Mareile, Simon Egbert, and Matthias Leese. "Predictive Policing and the Politics of Patterns." *The British Journal of Criminology* 59, no. 3 (May 2019): 674–692.

King, O.C., and Mertens, M. "Self-fulfilling Prophecy in Practical and Automated Prediction." *Ethics, Theory, and Moral Practice* 26 (2023): 127–152. https://doi.org/10.1007/s10677-022-10359-9.

Matsumi, Hideyuki, and Daniel J. Solove. "The Prediction Society: AI and the Problems of Forecasting the Future." *GWU Legal Studies Research Paper* No. 2023-58, *GWU Law School Public Law Research Paper* No. 2023-58 (2025): 2-62.

Reports and Working Papers

Albrecht, Eduardo. "Predictive Technologies in Conflict Prevention: Practical and Policy Considerations for the Multilateral System." *UNU-CPR Discussion Paper Series*, June 2023, 1–17.

Bjola, Corneliu. "Artificial Intelligence and Diplomatic Crisis Management: Addressing the 'Fog of War' Problem." *DigDiploROx Working Paper No. 6*, Oxford Digital Diplomacy Research Group, Oxford Department of International Development, University of Oxford, July 2022, 2–14.

Cement, Heidelberg. “The Israeli Exploitation of Palestinian Natural Resources: Part II”. *Who Profits*, November 2016. Accessed March 2025. https://www.whoprofits.org/publications/report/128?the-israeli-exploitation-of-palestinian-natural-resources-part-ii-heidelberg-cement.

Druet, Dirk. “Enhancing the Use of Digital Technology for Integrated Situational Awareness and Peacekeeping-Intelligence.” *Thematic Research Paper for the DPO Peacekeeping Technology Strategy*. Center for International Peace and Security Studies, McGill University, April 2021, 1–20.

Just Peace Labs. “Technology in Conflict: Conflict Sensitivity for the Tech Industry.” Just Peace Labs, 2020. Accessed March 2025. https://drive.google.com/file/d/1T78rYRWDoZZNzS4WVNR7_Ems96Tsa169/view

Soave, Tommaso. “From the Rule of Law to the Rule of Code? Predictive Data Analytics and the Changing Contours of UN Humanitarian Response.” Central European University (CEU), May 4, 2021, 1–21.

Other

Pauwels, Eleonore. “Artificial Intelligence and Data Capture Technologies in Violence and Conflict Prevention: Opportunities and Challenges for the International Community.” *Global Center on Cooperative Security*, 2022.

Appendix

Appendix A – Code and Data Repository

Saffron Ria Sadiq. *Conflict Prediction Models: ACLED, Amnesty International, and Human Rights Watch*. GitHub repository, 2025. Accessed March 2025. https://github.com/saffronsadiq/DH_Conflict_Prediction.

Appendix B – Conflict Keyword List

General Conflict Terms:

"conflict", "war", "violence", "clash", "unrest", "hostility", "insurgency", "uprising", "turmoil", "rebellion", "riot", "chaos", "combat", "fighting", "military", "warfare", "skirmish", "confrontation", "aggression", "armed conflict", "battle", "civil war", "sectarian violence", "terrorism", "extremism", "radicalization"

Political Instability and Governance Issues:

"coup", "authoritarianism", "dictatorship", "regime", "crackdown", "suppression", "oppression", "corruption", "sanctions", "martial law", "repression", "autocracy", "state violence", "political prisoners", "dissident", "government collapse", "failed state", "instability", "power struggle", "human rights abuses", "surveillance", "police brutality", "extrajudicial killing"

Ethnic, Religious, and Sectarian Conflict:

"ethnic cleansing", "genocide", "sectarian", "ethnic violence", "religious persecution", "massacre", "pogrom", "xenophobia", "hate crime", "discrimination", "racial violence", "forced displacement"

Humanitarian Crises and War Crimes:

"refugee", "displacement", "asylum", "forced migration", "internally displaced persons", "IDPs", "famine", "blockade", "humanitarian crisis", "siege", "war crimes", "crimes against humanity", "atrocity", "ethnic persecution", "rape as a weapon of war", "child soldiers", "torture", "execution", "disappearances", "chemical weapons", "biological weapons", "cluster munitions", "landmines", "indiscriminate attacks"

Terrorism & Insurgency

"terrorist", "suicide bombing", "extremist", "radical group", "paramilitary", "guerrilla warfare", "fundamentalism", "insurgent", "armed group", "warlord", "militant", "jihad", "proxy war"

Military Operations and Foreign Interventions:

"airstrike", "bombing", "drone strike", "invasion", "occupation", "military intervention", "proxy conflict", "foreign interference", "sanctions", "arms trade", "weapons proliferation"