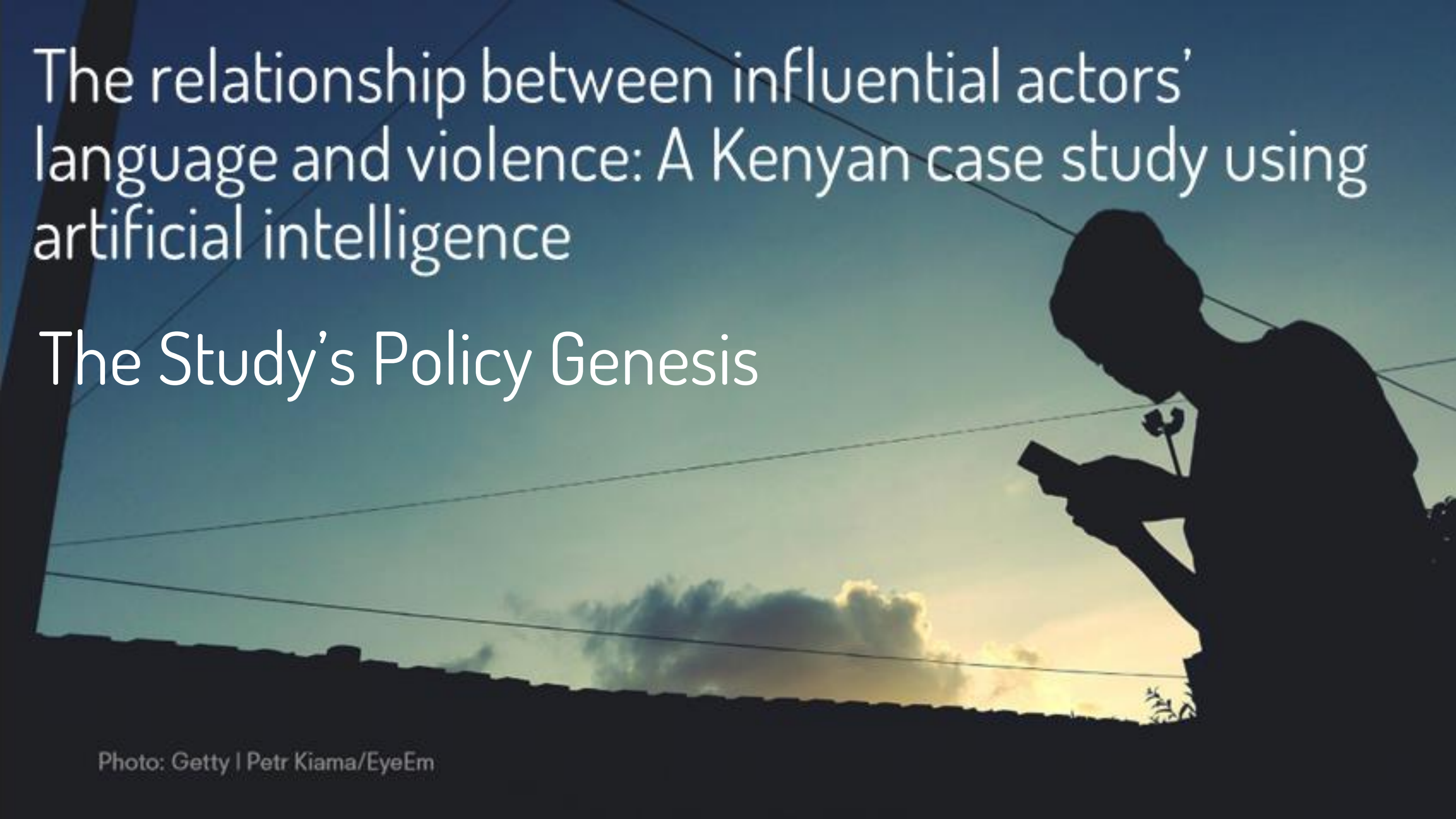
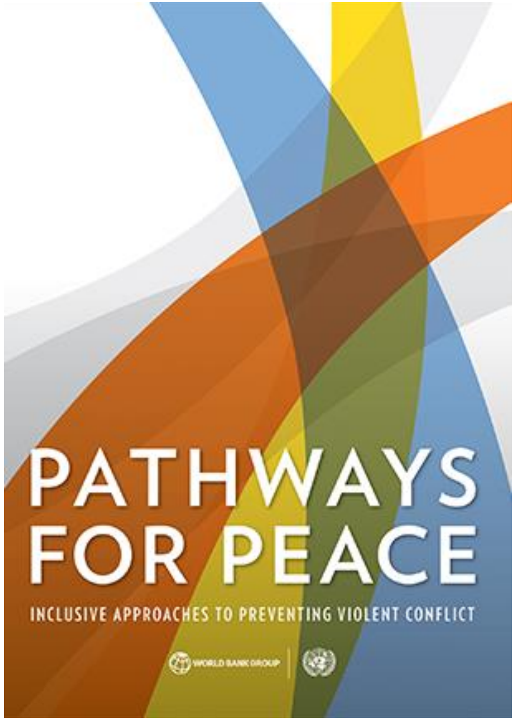

The relationship between influential actors' language and violence: A Kenyan case study using artificial intelligence
The Study's Policy Genesis
Photo: Getty | Petr Kiama/EyeEm

Imperative for this Study Emerges from: Joint UN-World Bank flagship study on Conflict Prevention (Pathwaysforpeace.org)

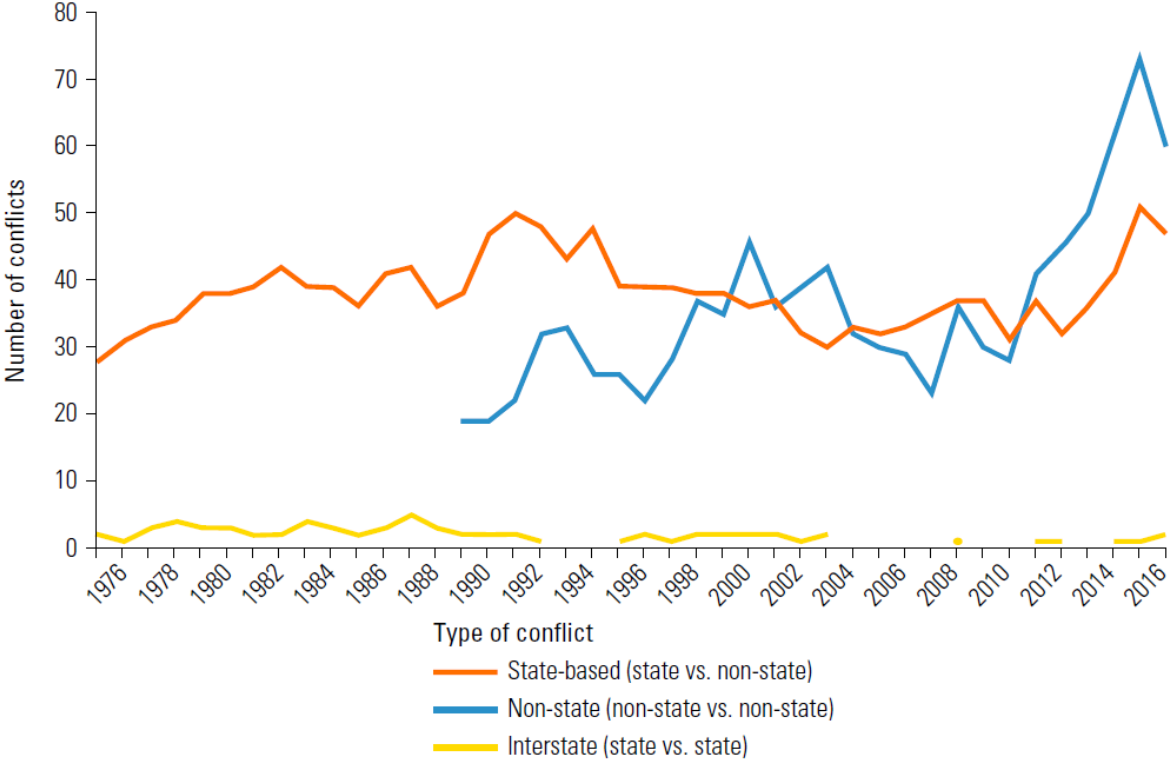
- **United Nations / World Bank Findings (2018):**
  - **Development contributions to prevention should be risk-informed** and account for multidimensional nature of risk in an increasingly interdependent world.
  - Addressing underlying risks of conflict should be targeted towards **attending to grievances based on real and perceived group exclusions** and counteracting ways that can be used to mobilize for violence.
  - **Most successful prevention is endogenous**, undertaken by local or national actors. International actors can support these broad and inclusive processes. In this sense, prevention enhances sovereignty.
  - **Effective prevention today is built on broad coalitions.** National governments hold the primary responsibility for prevention, but need to include civil society, private sector and regional actors.



- The declining trend of conflicts worldwide reversed in 2010

- Battle-related deaths, number of armed conflicts, civilian casualties, terrorist attacks, number of refugees and violently displaced people have all increased

FIGURE 1 Conflict Trends



The Business Case for Prevention

- Prevention saves lives, is cost-effective and avoids immense economic losses that accompany conflicts

→ It is cost-effective!

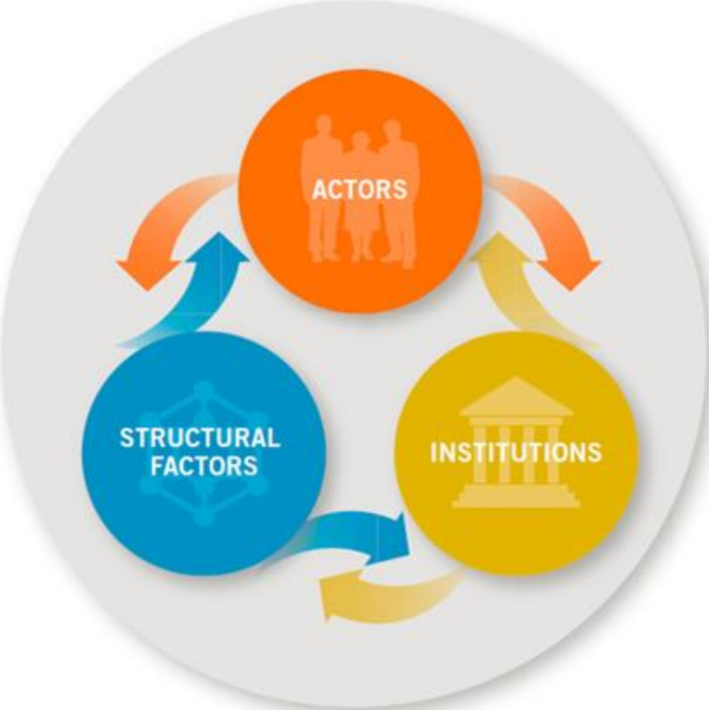
TABLE 1 Modeling the Returns on Prevention under Three Scenarios

	Scenarios		
	Optimistic	**Neutral**	**Pessimistic**
Assumptions:			
Lost GDP growth per conflict year (percent)	5.2	3.9	2.5
Cost of prevention (US$, millions)	100	500	1,000
Effectiveness of prevention (percent)	75	50	25
Prevented damage (US$, millions)	68,736	34,251	9,377
Saved costs (US$, millions)	1,523	1,176	698
Additional cost (US$, millions)	–352	–2,118	–5,247
Net savings per year (US$, millions)	**69,907**	**33,309**	**4,828**

Sources: Collier and Hoeffler 2006, Dunne 2012, Lomborg 2013.
Note: Prevented damage is the prevented economic damage and deaths; *saved costs* are the saved costs from late intervention costs related to peacekeeping and humanitarian assistance that become unnecessary with prevention; and *additional costs* are the additional costs needed for prevention efforts.

FIGURE 4 Three Core Elements of the Pathway

ACTORS are leaders or social groups who make decisions, in competition or cooperation with one another, that determine how a society moves forward.



STRUCTURAL FACTORS are elements that are hard to change except in the long term and that affect the fundamental nature of the sociopolitical system.

INSTITUTIONS comprise political and institutional rules that can change only in the medium term, and that strongly shape actors' behavior, incentives, and capacity to work together.

UN Management Response – Sample Recommendations: Risk Management

- Underlying risks of conflict pertain to **objective** and **perceived political, economic and social exclusions** of groups. They should be monitored in four key arenas of contestation: **political power; security and justice, natural resources and services.**
- International partners should commit to support development of **national and regional platforms** to monitor and collectively mitigate risks.
- This should include efforts **to harness technology** through data sciences and integrating perceptions monitoring into data collection and conflict analyses.
- **Multilateral partners should share risk assessments** across humanitarian, development, peace and security actors. Risk monitoring systems should be **linked with cross sectoral capacities to act**, inform decision making and operational guidance.
- **So how do we determine objective risk posed by influential actors?**

FIGURE 5 Siloed Approach to Prevention

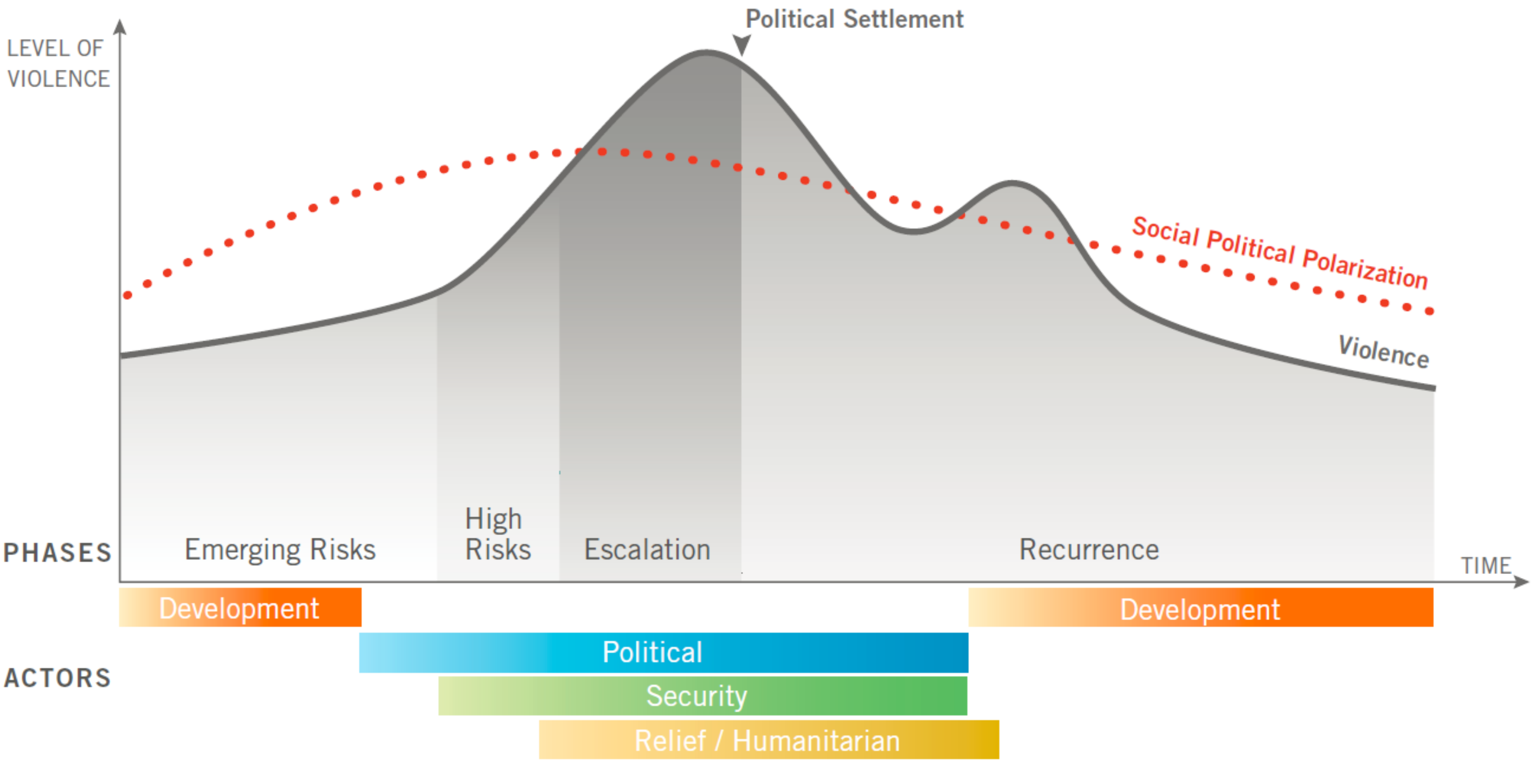
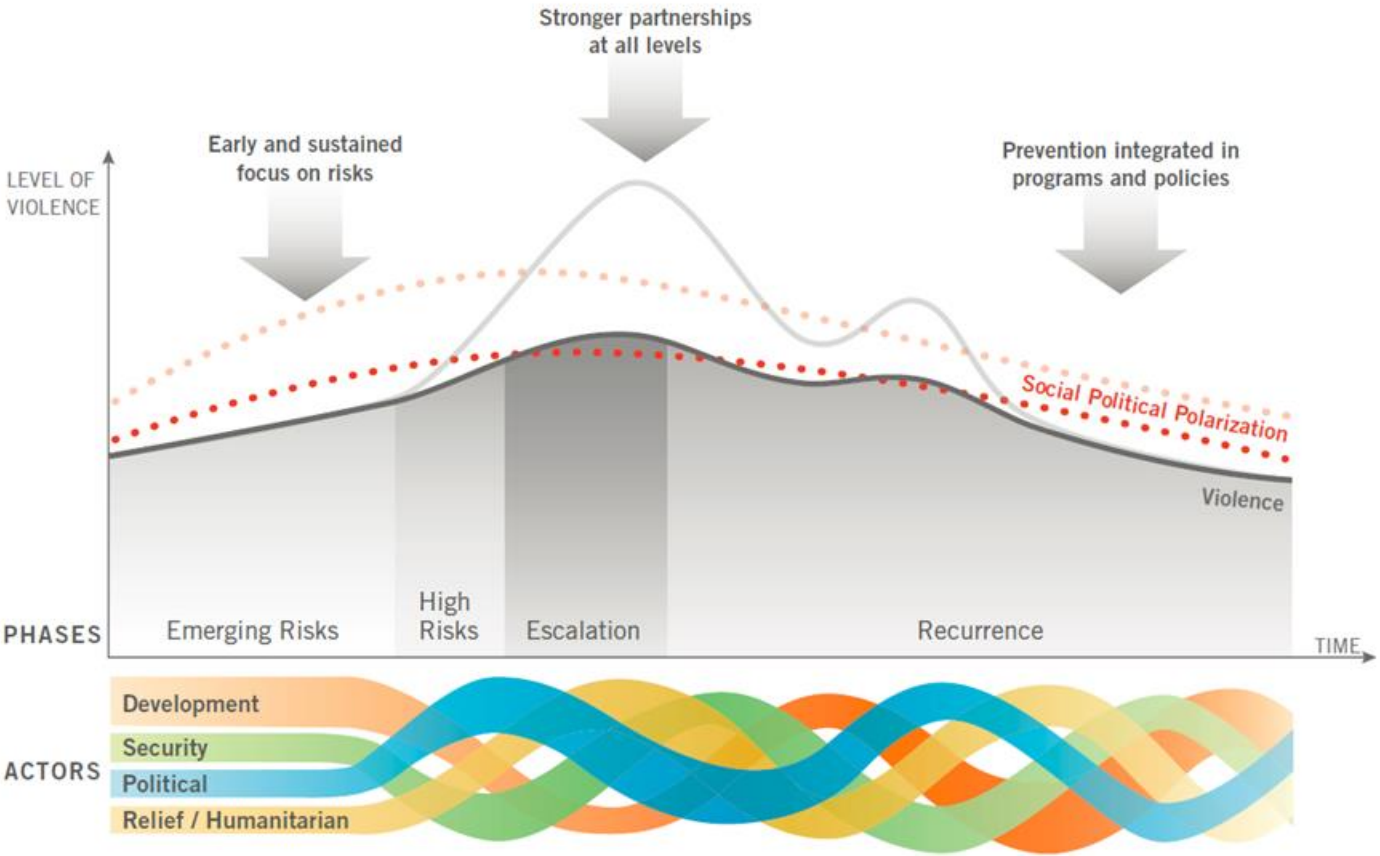


FIGURE 6 Integrated Approach to Prevention



Leaders' intent: a risk assessment knowledge gap

- **Risk-informed** and account for multidimensional nature of risk
  - That shocks (such as natural disasters, commodity price change, or other economic shocks) interact with the risk posed by grievances based on real and perceived group exclusions
- How to counteract scope for mobilization to violence when shocks interact with risks?
  - Leaders mobilize people towards or away from violence – how can societies positively shape their intent?
  - Identifying increased risk associated with behavior empowers stakeholders - *the evidence of intent is in the language.*
  - Local or national actors undertake successful prevention

Identifying leaders' intent via language

- A scientific basis for observing the risk residing in leaders' language is necessary to empower local and external actors to cite, without subjectivity, the risk of violence their language is associated with.

- First problem: Identify the leaders. Engage academics (across social science) and informed observers specializing in a country to develop a list of political, social, security and economic leaders that may mobilize people towards and away from violence.

- Second problem, identify their language, how it changes, the sentiment associated with particular leaders, and how it relates to violence.

Examine ‘influential’ actors’ language, not broader discourse

- A lot of work considering the relationship between the broader populations’ discourse and violence.
- This thesis considers the inclination of powerful individuals *to mobilize* people to violence, not the vulnerability of the population or social groups *to be mobilized*.
- Therefore, we test powerful actors’ own speech as a proxy for their intent to mobilize people towards or away from violence.
- We do that by testing the relationship of their language with changes in the levels of violence.

This is the first step – the evidence of the relationship

- We are not asserting causation, but rather demonstrating that, scientifically, there is a relationship.
- Lurking variables may exist, but observers will be endowed with a metric for objectively identifying speech that increases or decreases risk.
- This work takes us towards the point at which observers may say:

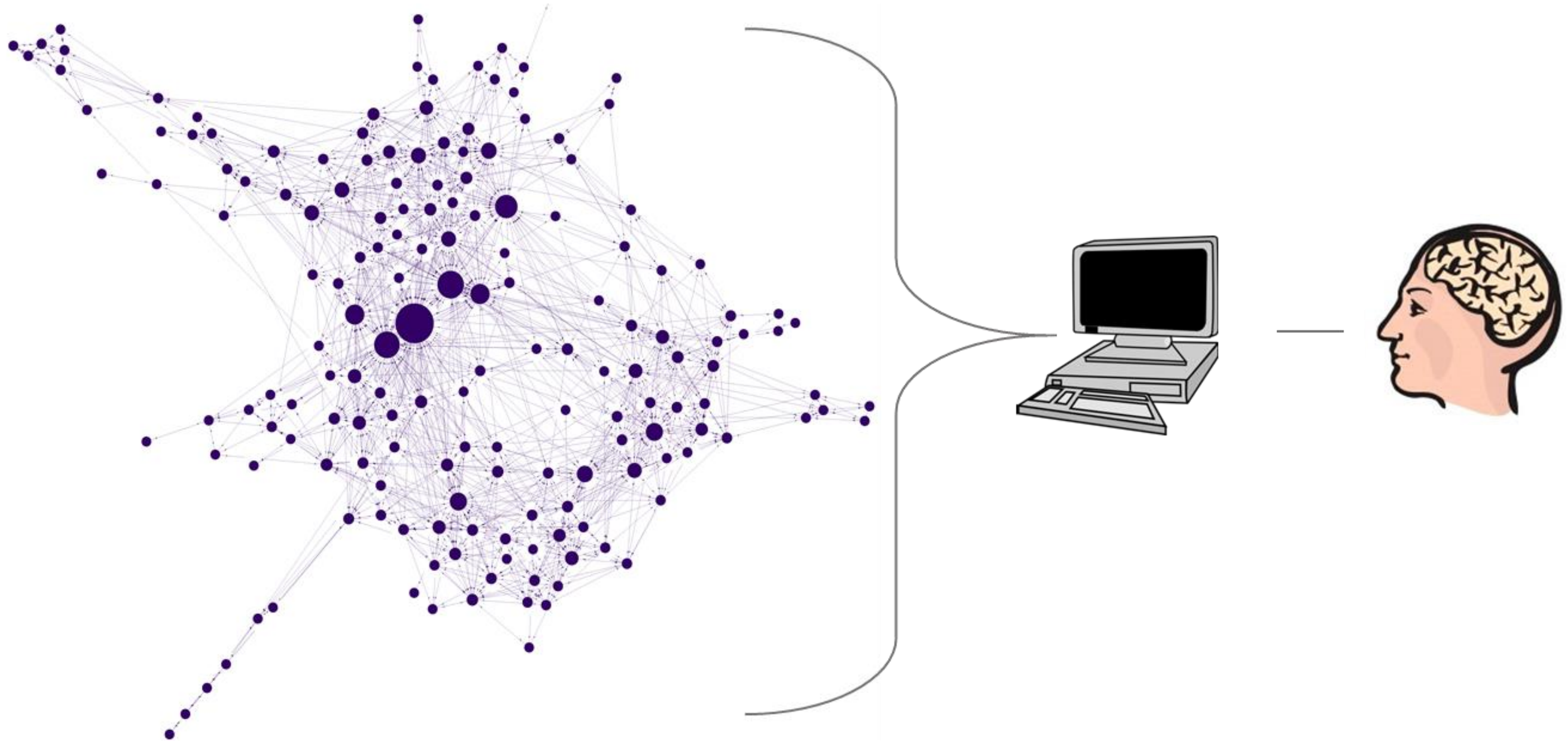
"Whether you intend to incite violence or not, your language is associated with an increased risk."

The relationship between influential actors' language and violence: A Kenyan case study using artificial intelligence

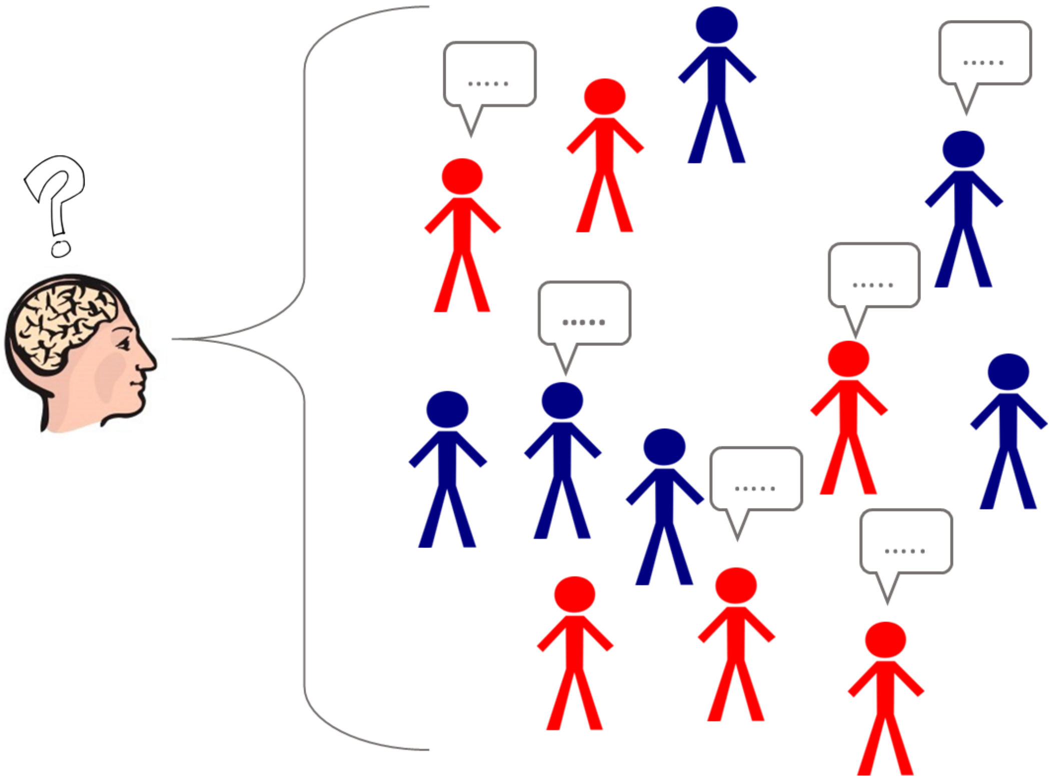
Findings

Photo: Getty | Petr Kiama/EyeEm

Sentiment Analysis technology allows us to monitor large quantities of online data.

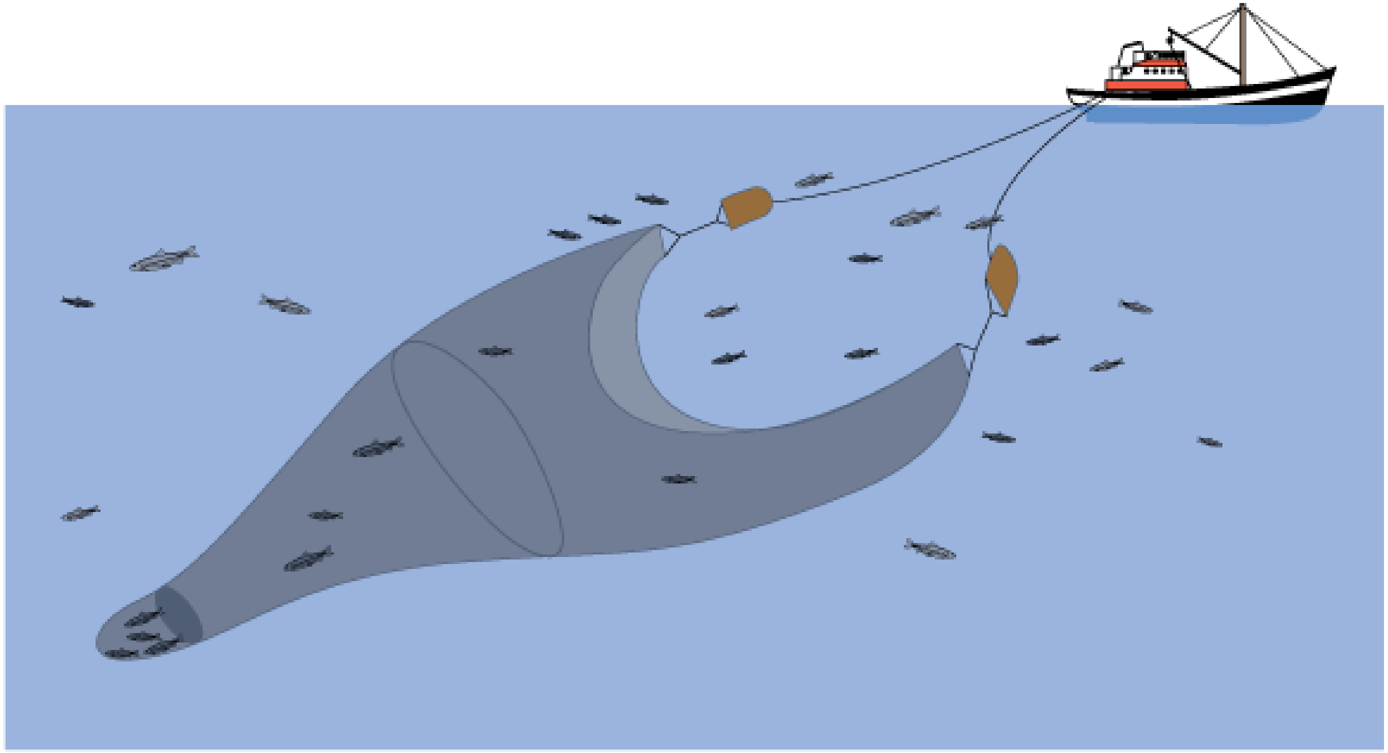


Including data concerning online speech.



Research Question: Can we use this technology to identify correlations between online speech and political violence in Kenya?

We collected sentiment scores for 6100 speech instances from Twitter.



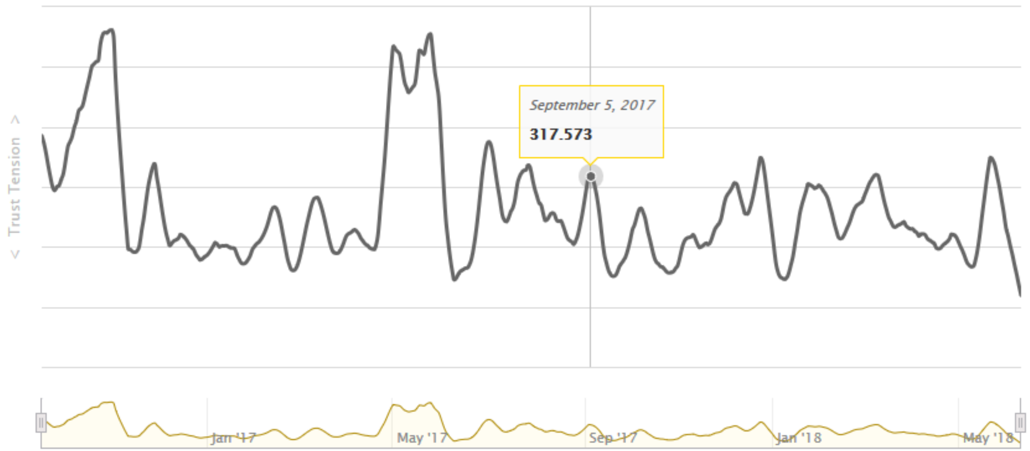
Software reads and understands the meaning of words in multiple languages, and allocates sentiment scores and time reference.

"Anti-government protests expected next Tuesday in Nairobi"

Negative Sentiment

Predictive

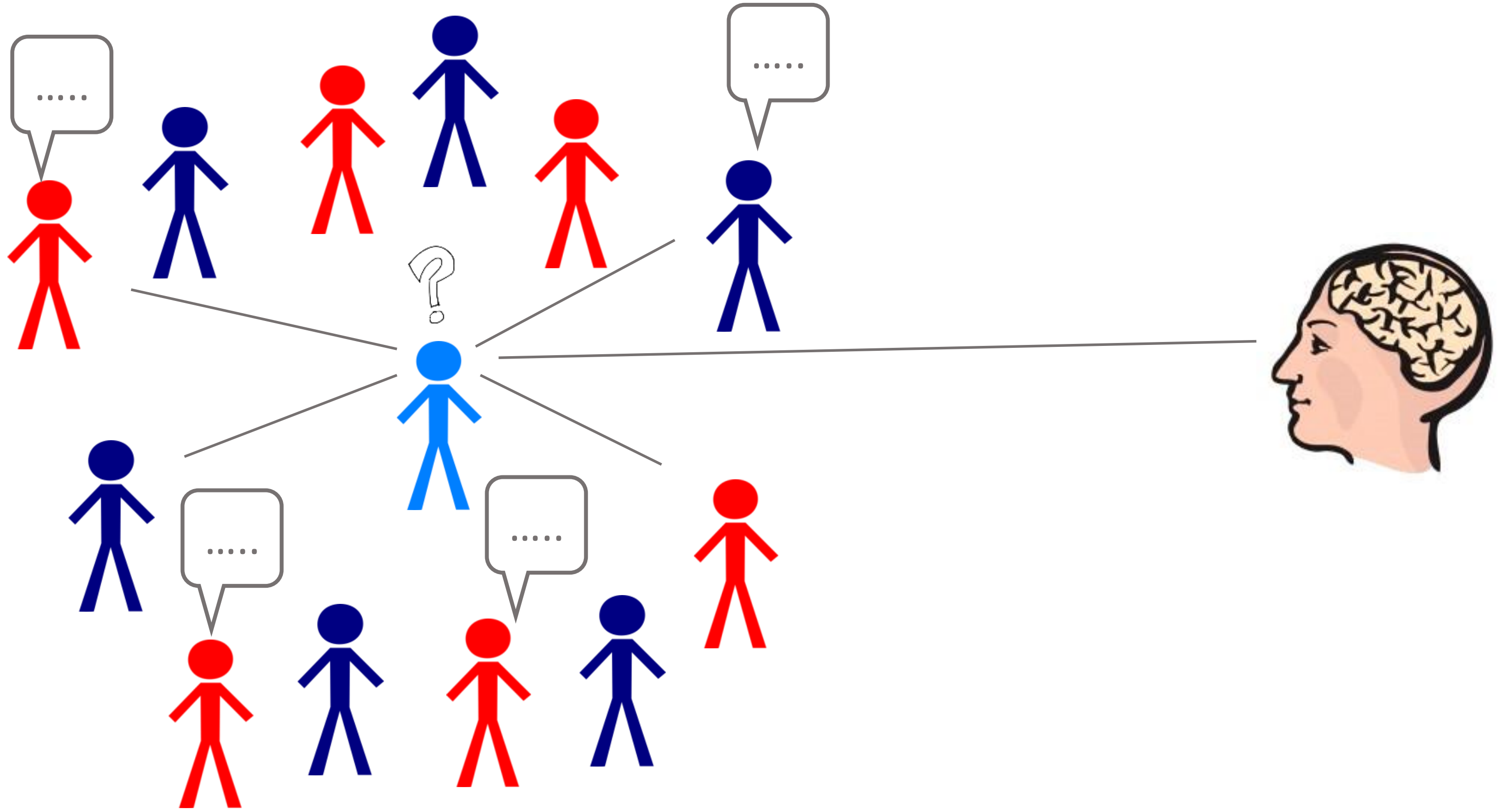
Unstructured textual data (i.e. a tweet) is transformed into structured data (i.e. a number) to gauge the changing emotional tone over time.



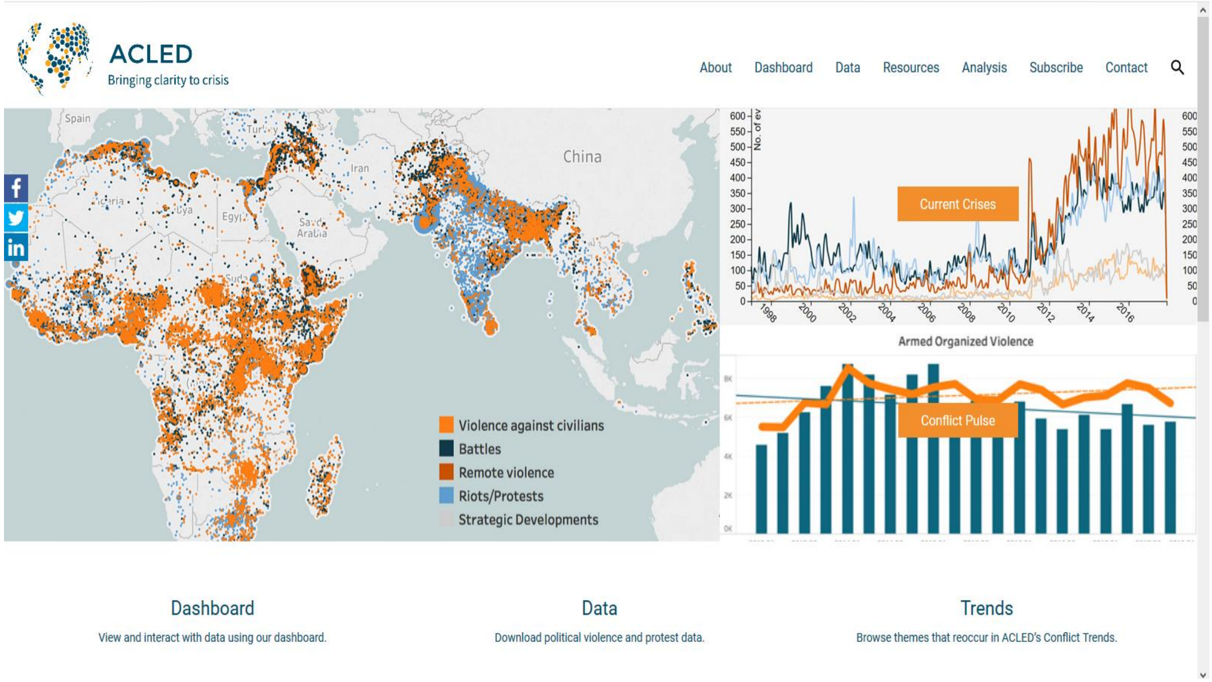
Daily sentiment scores were collected for 30 key political actors in Kenya from January 2012 to December 2017.

	A	BQU	BQV	BQW	BQX	BQY	BQZ	BRA	BRB	BRC	BRD	BRE	BRF	BRG
1	Topic: South Sudan	Apr 18 2017	Apr 19 2017	Apr 20 2017	Apr 21 2017	Apr 22 2017	Apr 23 2017	Apr 24 2017	Apr 25 2017	Apr 26 2017	Apr 27 2017	Apr 28 2017	Apr 29 2017	Apr 30 2017
2	Blog Predictive Instances	3658	3355	2989	2243	511	771	2312	2056	1922	2485	2289	1219	1480
3	Blog Predictive Positive	1636	1335	822	807	173	340	932	671	425	687	748	510	515
4	Blog Predictive Negative	218	150	396	319	60	30	85	204	261	136	54	38	103
5	Blog Predictive Neutral	1804	1870	1771	1117	278	401	1295	1181	1236	1662	1487	671	862
6	Government Predictive Instances	66	41	40	64	1	6	45	93	74	7	2	2	4
7	Government Predictive Positive	5	6	3	24	1	0	1	66	16	1	1	0	1
8	Government Predictive Negative	1	2	2	0	0	1	0	0	3	0	0	0	0
9	Government Predictive Neutral	60	33	35	40	0	5	44	27	55	6	1	2	3
10	Mainstream News Predictive Instances	6514	6887	7176	5818	2364	2255	8384	5600	5886	5714	5316	2370	2079
11	Mainstream News Predictive Positive	3168	3336	3488	2849	1065	1012	3343	2993	3199	3193	2336	1234	1093
12	Mainstream News Predictive Negative	470	439	560	382	179	246	569	438	417	464	515	260	253
13	Mainstream News Predictive Neutral	2876	3112	3128	2587	1120	997	4472	2169	2270	2057	2465	876	733
14	Social Media Predictive Instances	208	338	242	264	166	116	140	216	202	166	170	137	144
15	Social Media Predictive Positive	28	61	41	68	19	18	28	66	36	21	32	31	26
16	Social Media Predictive Negative	4	4	18	32	19	6	6	20	27	20	9	20	13
17	Social Media Predictive Neutral	176	273	183	164	128	92	106	130	139	125	129	86	105
18	Blog Not Predictive Instances	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
19	Blog Not Predictive Positive	4318	4623	4301	4165	3857	4413	4645	4581	4393	4646	4520	4468	4168
20	Blog Not Predictive Negative	1505	1290	1251	1204	1219	1081	1197	1177	1308	1189	999	1036	1034
21	Blog Not Predictive Neutral	4176	4086	4447	4630	4923	4505	4157	4241	4298	4164	4480	4495	4797
22	Government Not Predictive Instances	3570	4725	3041	3437	407	330	2349	4150	2344	366	468	245	249
23	Government Not Predictive Positive	208	757	294	354	40	39	234	461	385	37	85	61	23
24	Government Not Predictive Negative	33	205	49	32	9	14	48	40	43	15	25	10	4
25	Government Not Predictive Neutral	3329	3763	2698	3051	358	277	2067	3649	1916	314	358	174	222
26	Mainstream News Not Predictive Instances	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
27	Mainstream News Not Predictive Positive	4965	4922	4616	4765	4986	5143	5086	5214	4735	5063	4615	4261	4234
28	Mainstream News Not Predictive Negative	932	918	1227	1165	1058	909	1031	862	1081	906	1225	1240	1598
29	Mainstream News Not Predictive Neutral	4102	4159	4156	4069	3955	3947	3882	3923	4183	4030	4159	4498	4167
30	Social Media Not Predictive Instances	9999	9999	9999	9999	9999	7196	8965	9999	9999	9999	9999	8116	9999
31	Social Media Not Predictive Positive	1605	1495	1420	1538	1305	1278	1784	1603	1933	1800	2175	1224	1460
32	Social Media Not Predictive Negative	1044	926	1113	1408	1306	1092	1294	1973	1506	1993	1686	1523	1519
33	Social Media Not Predictive Neutral	7350	7578	7466	7053	7388	4826	5887	6423	6560	6206	6138	5369	7020

Actors were identified through a network of experts on Kenyan political dynamics with extensive local knowledge.



The variable for political violence is the daily number of fatalities as reported by the Armed Conflict Location and Event Data Project (ACLED).



We divided the data into two sets: train and test

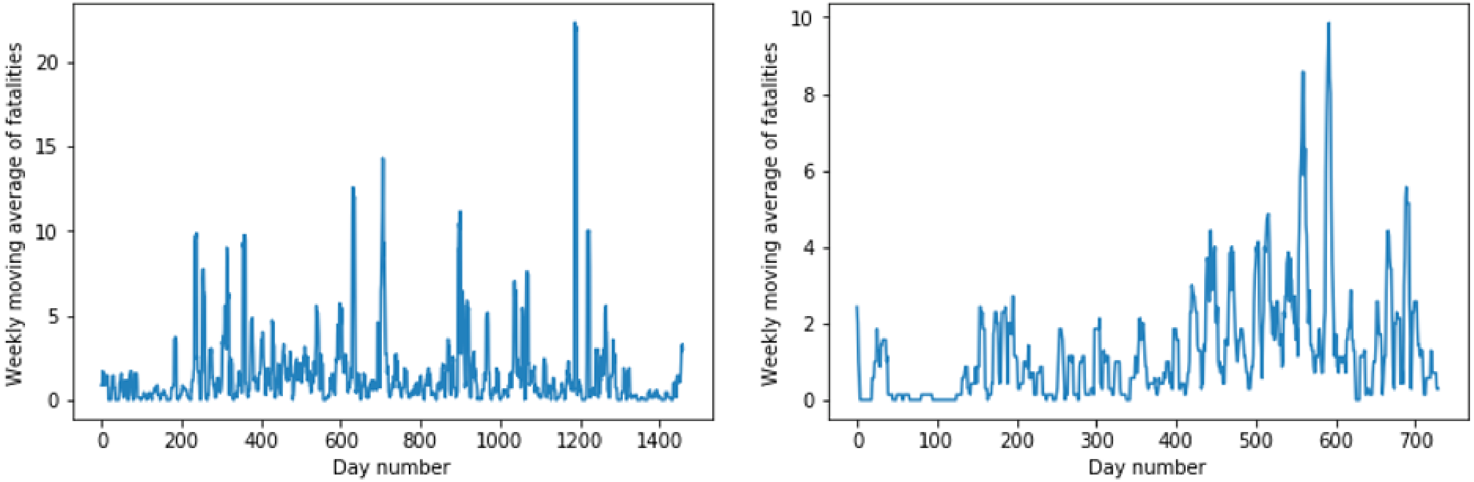


Figure 1: Change in weekly moving average of fatalities in Kenya for the periods 2012-2015 (left) and 2016-2017 (right).

The training set is fed into a Random Forest classifier model:

On each day we feed data for the last 30 days.

Ask the model to predict what the average fatalities of the timeframe within the next N days will be.*

The categories are two: higher or lower/in-range relative to the overall average fatalities of the entire data set.

Higher is 25% higher than overall average, otherwise it is considered lower/in-range.

*N is a variable that refers to the 'look ahead' period of our predictions. For instance, if the look ahead period is 60 days, the sentiment data belonging to the last 30 days will be used to predict what the average fatalities within the next 60 days will be.

The model is then used on the test set using different look ahead time frames in order to explore performance.

Predictive accuracy nears 85% when the look ahead period is between 100 and 125 days.

To evaluate the performance, we use the ROC Area Under Curve (AUC) metric, which is used to confidently measure success of machine learning models.

ROC AUC becomes 0.5 as the model makes random or uninformative predictions as it nears 1.0 as it makes correct predictions.

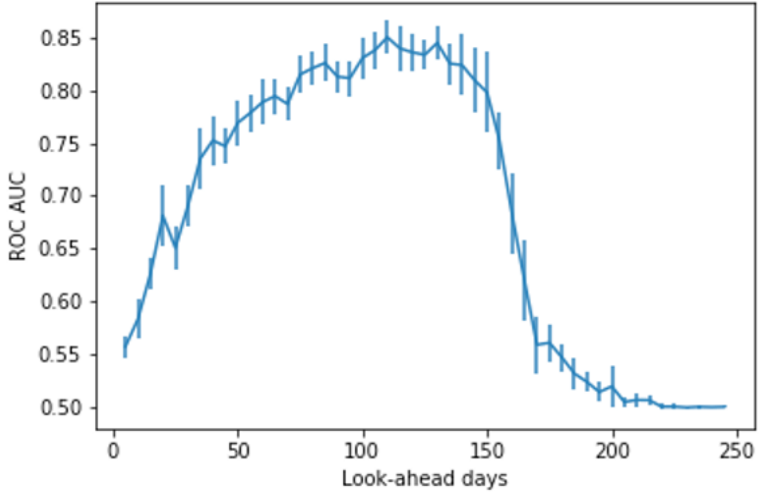


Figure 2: Performance of the model for the test set at different look ahead days.

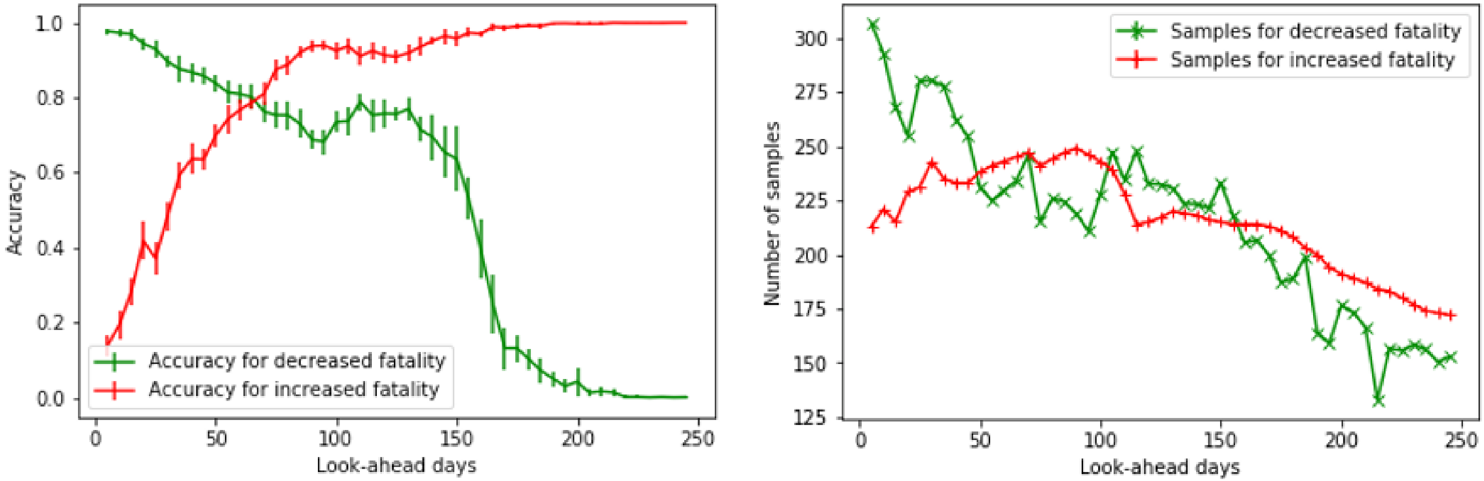


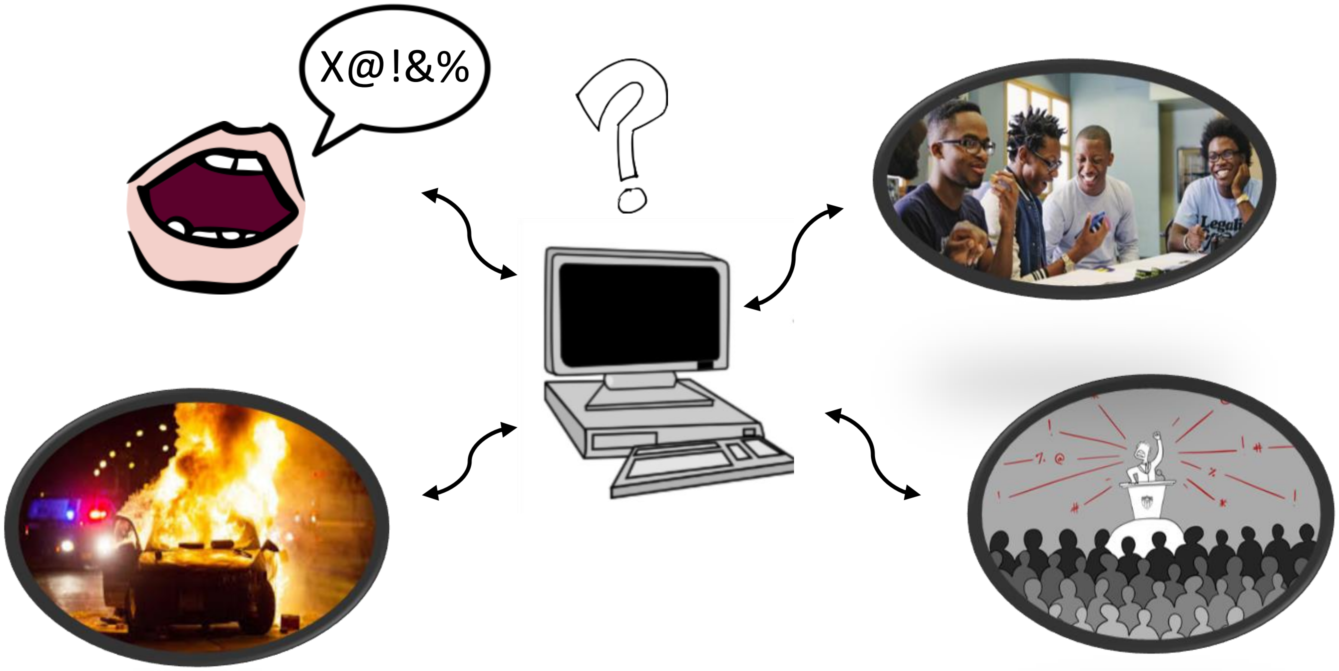
Figure 3: Prediction accuracy for each category (at left) and the number of samples (at right) for the test set.

The relationship between influential actors' language and violence: A Kenyan case study using artificial intelligence

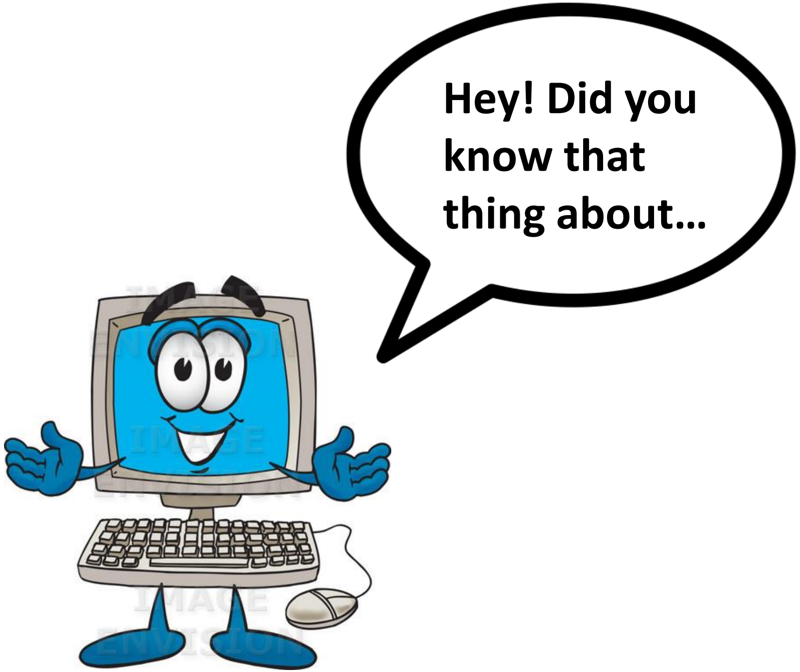
Future directions of research

Photo: Getty | Petr Kiama/EyeEm

As we move forward we plan to further develop this software so that it can **adapt to changing discursive/sociopolitical environments**.

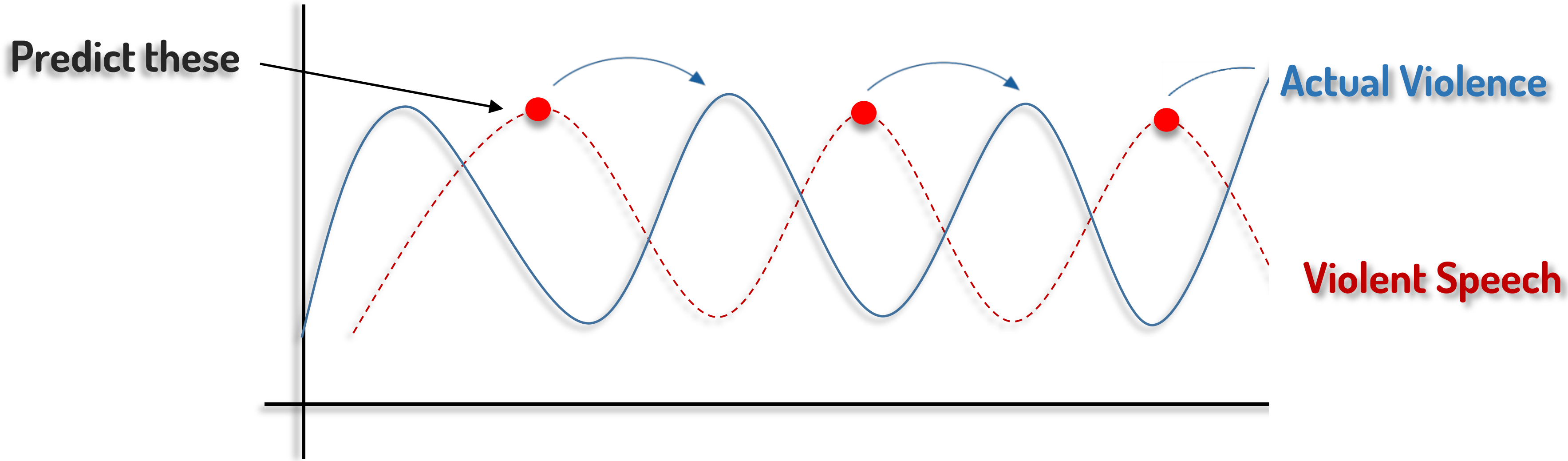


We plan to explore the design of an app that has the potential to autonomously adapt to changes in context – and **to offer explanatory models** for observed phenomena.

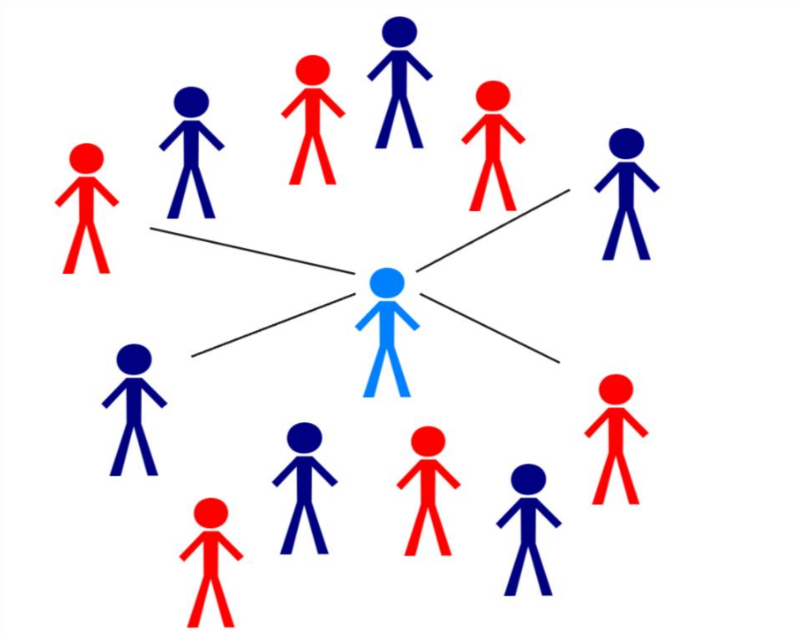


We envision software that can offer explanations as to which **specific individuals' speech** is most associated with increased risk of violence, and, assess those individuals' contributions to risk using a heuristic framework based in legal definitions of incitement.

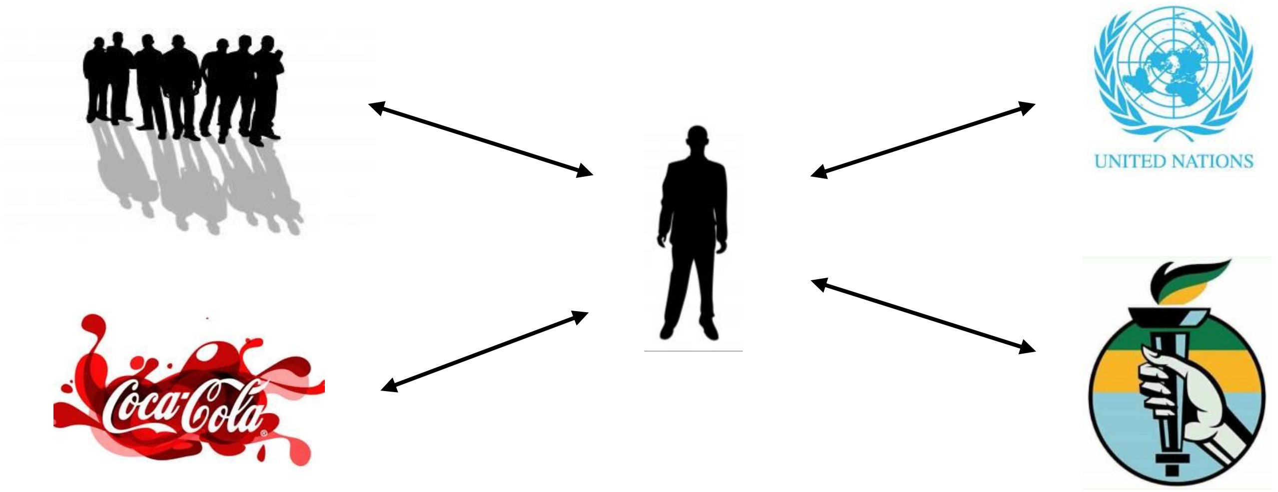
The app would identify these as “speech events of elevated risk” and run a secondary analysis designed to predict this future language by specific persons associated with increased risk of violence.



To predict patterns in cognition we posit that it is necessary to collect adequate amounts of contextual and relationship data. We acknowledge the complexity and context specificity involved in identifying, coding, and testing such data.

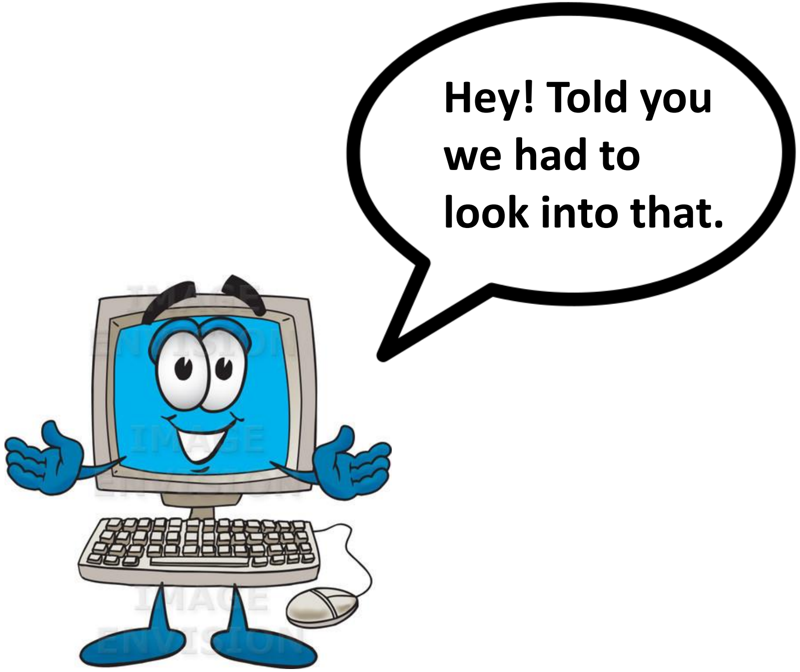


To this end, the app would run a third analysis designed to identify additional contextual dynamics (i.e. **identifying linkages, and patterns of linkages**, to other influential individuals, national and multinational corporations, foreign actors, social groups, and national and international organizations).



The app automatically adds data feeds for these additional actors/entities, and then runs the previous analyses again until peak predictive accuracy is reached.

Such a capacity may contribute to the objective of: “shaping a future in which AI-enabled machines serve as trusted, collaborative partners in solving problems of importance to national security.”



To achieve this we propose a layered AI architecture where multiple machines interact:

Machine A) assesses increased risk of conflict in a context.

Machine B) identifies specific individuals' utterances associated with increased risk.

Machine C) assesses these utterances in terms of the legal framework for incitement.*

Machine D) assesses increased risk of specific individuals' incitement speech events.

Machine E) identifies linkages concerning contextual dynamics.

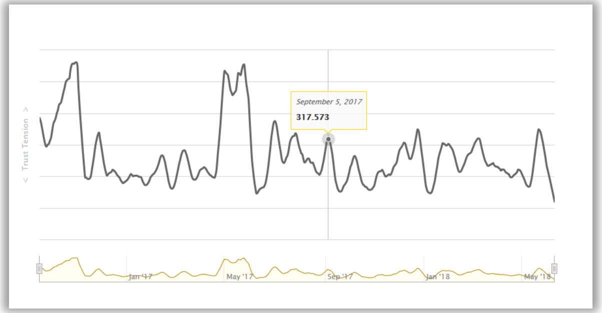
Machine F) automatically harvests new data feeds.

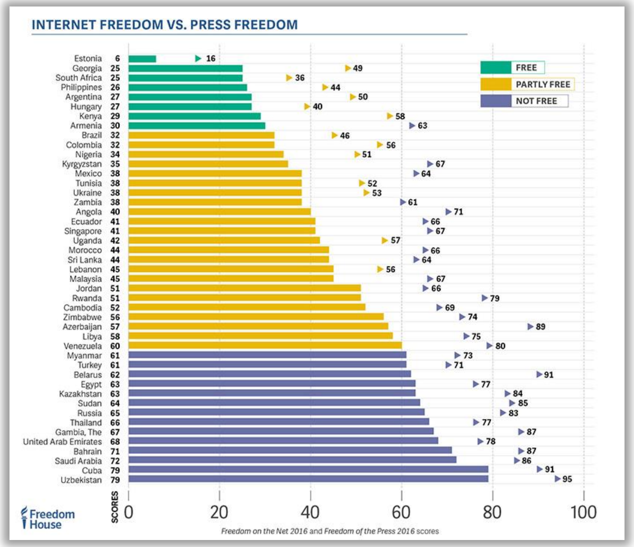
Machine A) process repeats until accuracy no longer increases.

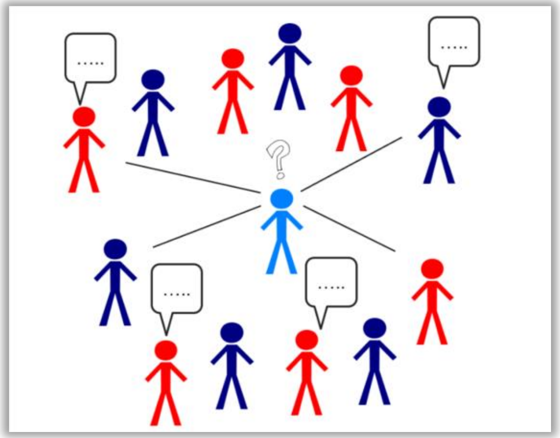
* This framework will be explained below.

There are three types of data the software will rely upon.

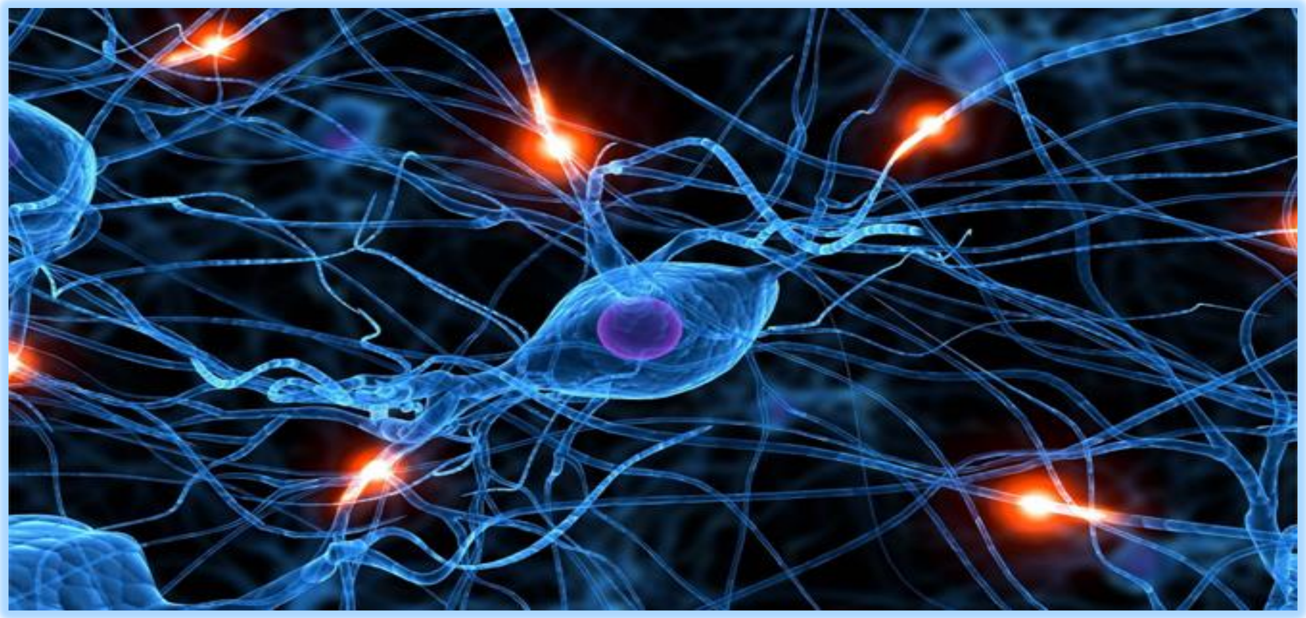
Sentiment scores, econometric and political statistics, and **relationships data**.



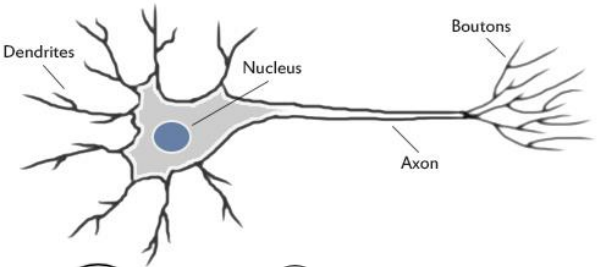


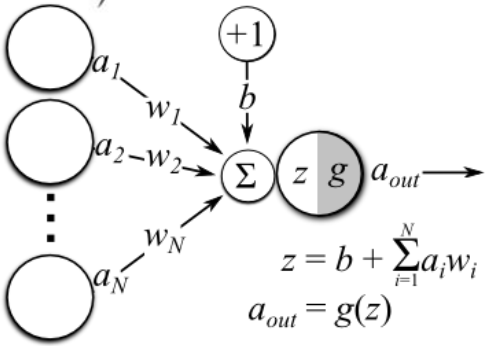


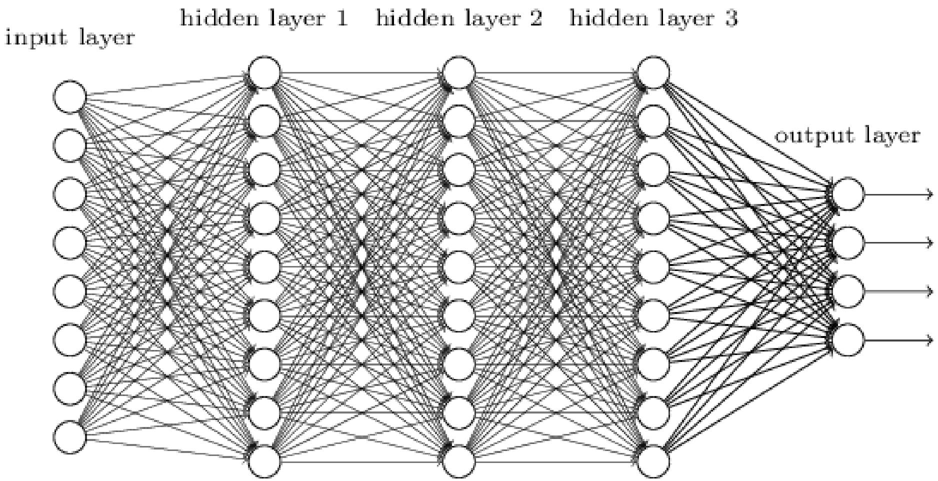
To conduct this research we intend to use deep neural networks and to do that sufficient amounts of data must be collected.



An artificial neural network consists of a network of simple information processing units, called **neurons**, just like a human brain. We build deep learning models by creating many neurons and linking them to construct layered architectures.







Sentiment scores would be gathered via a variety of online and offline sources (i.e. public **social media** accounts, mainstream media **news articles** and **comments** sections, **blog posts**, and **public radio** talk show dialogue converted into text).

These scores are further divided according to utterances made **by** a specific influential actor, utterances made **about** him or her, and organized by whether they reference **future events or not**.

Radios - especially in vernacular languages - remains the most important vehicle of potentially mobilizing speech.

Chem chemi ya Ukweli

RADIO CITIZEN

wadr

West Africa Democracy Radio

ON AIR IMPACT 103FM

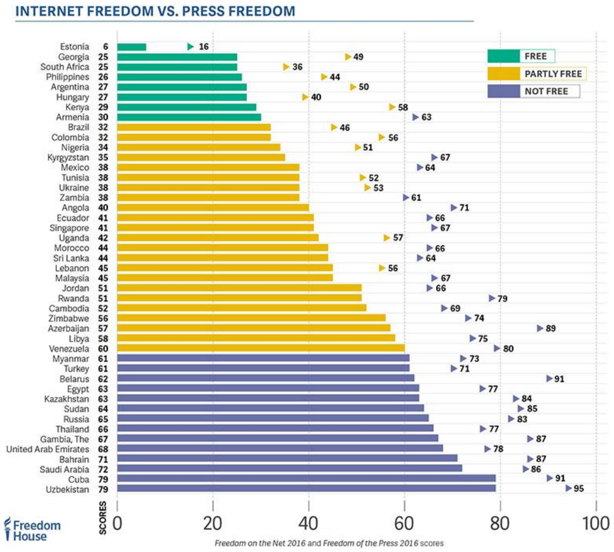
SABC

We plan to address this by including a capacity that transcribes radio speech to text and an NLP capacity that can attribute sentiment in local vernacular languages.

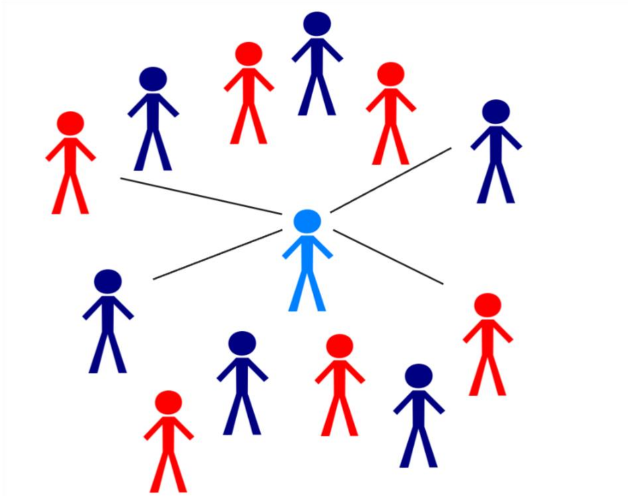
Econometric data includes fluctuations in food, fuel and other commodity prices, currency exchange rates, stock market indices.

Political statistics measure variations in levels of:

- Electoral credibility
- Press freedom
- Civil society robustness
- Police deployment
- Restrictions on fundamental freedoms
- Access to justice, education, livelihoods and healthcare
- Attendance to public protests and rallies.



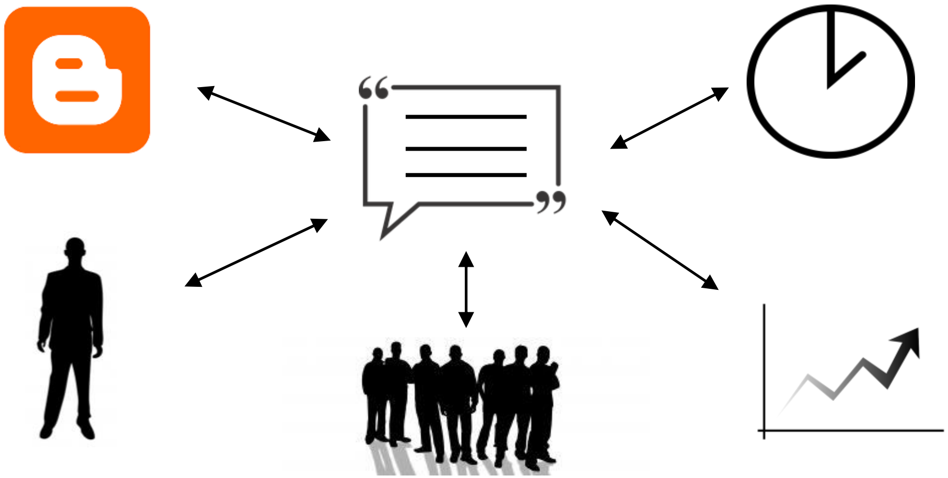
Relationship data is gathered through measuring the frequency of co-mentions of individuals to other individuals, corporations, organizations, social groups, etc. in the same variety of online and offline sources harvested for sentiment data.



Relationship data is key in allowing the software to measure gradations of speech risk.

By overlaying these three data types the software may embed an utterance within a matrix that accounts for the following factors associated with incitement to violence when testing the relationship to violence:

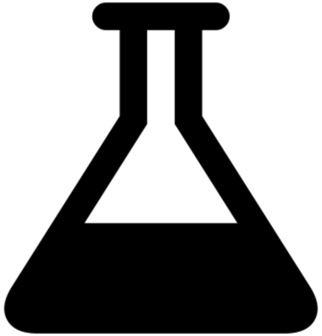
a) the content of the message
b) the type of medium
c) the time gap between when a message is said and events referenced
d) the economic and political context
e) the identity of the originator of the message
f) the relationship between originator and audiences



(Theory of 'incitement' – Gregory Gordon 2017)

It is important to accounting for the actual contribution of speech to risk. To control for this we should run the experiment with and without sentiment data, thereby measuring the relative impact of actual utterances.

Experiment I



Experiment II

We also should consider the ethical implications of this research and any resulting predictive capacity.

There is a legitimate concern regarding freedom of speech and the relationship of speech to violence. Balancing the right to freedom of speech and the right to security of the person is difficult.

Finally, our previous review of literature focused on Anglo-American studies using AI to consider speech and violence (variant forms of violence and discourse).

What is the scientific discourse in, for example, Chinese, Nigerian, Indian, and Russian academia on the topic?

This literature can inform incorporation of directions of inquiry.

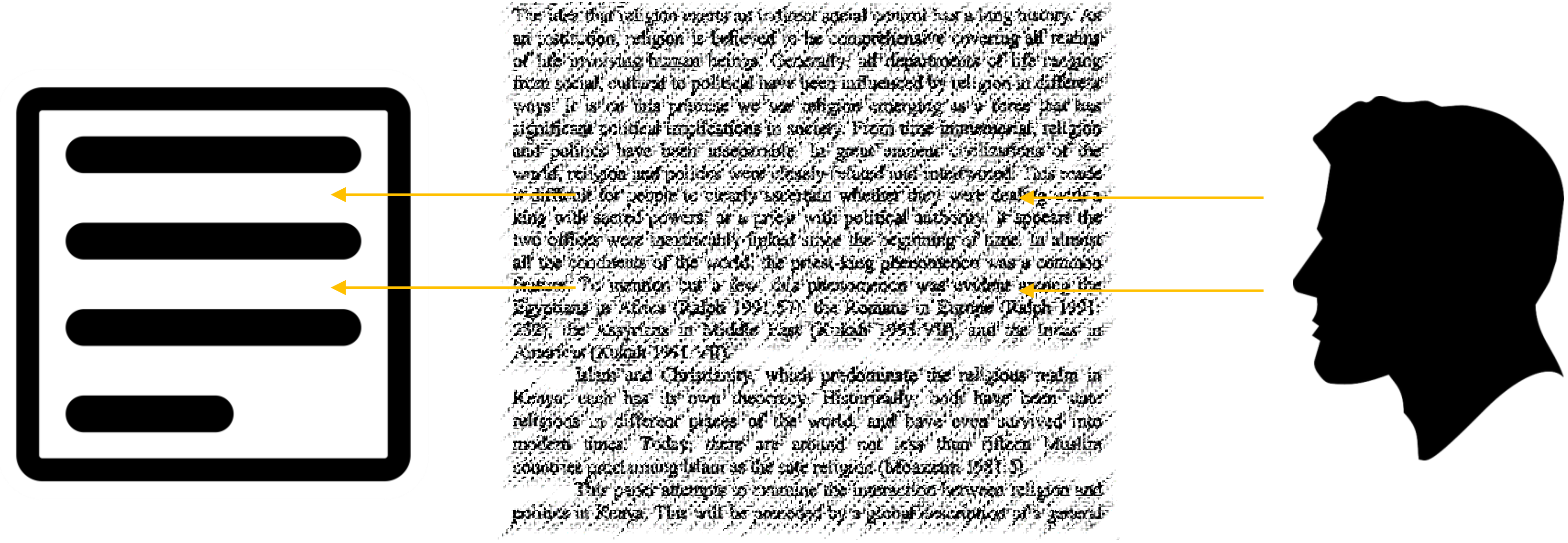

The relationship between influential actors' language and violence: A Kenyan case study using artificial intelligence
Navigating Ethical Dilemmas in an Emerging Field
Photo: Getty | Petr Kiama/EyeEm

Evidence of a relationship between language and violence may lead in the direction of empowering organizations to hold influential actors accountable for their speech or influence.

Yet, there is a risk that when early warning signals generated by algorithmic analysis are brought to government officers, governments can become overly sensitive and push back on any information that challenges their legitimacy or sovereignty.

A reflection on "packaging" of evidence for best policy application will eventually be needed.

Another issue is the fact that high profile political actors are learning to pay attention to what they say publicly, particularly in countries where the ICC has been involved, and thus speak in a coded way (often in different dialects) to convey messages between the lines.



The idea that religion exerts an indirect social control has a long history. As an institution, religion is believed to be comprehensive covering all realms of life involving human beings. Generally, all departments of life ranging from social, cultural to political have been influenced by religion in different ways. It is on this premise we see religion emerging as a force that has significant political implications in society. From time immemorial, religion and politics have been inseparable. In great ancient civilizations of the world, religion and politics were closely related and intertwined. This made it difficult for people to clearly ascertain whether they were dealing with a king with sacred powers, or a priest with political authority. It appears the two offices were inextricably linked since the beginning of time. In almost all the continents of the world, the priest-king phenomenon was a common feature. To mention but a few, this phenomenon was evident among the Egyptians in Africa (Ralph 1991:57), the Romans in Europe (Ralph 1991: 252), the Assyrians in Middle East (Kulah 1955:[illegible]), and the Incas in America (Kulah 1951:[illegible]).

Islam and Christianity, which predominate the religious realm in Kenya, each has its own theocracy. Historically, both have been state religions in different places of the world, and have even survived into modern times. Today, there are around not less than fifteen Muslim countries proclaiming Islam as the sole religion (Moazzam 1981:5).

This paper attempts to examine the interaction between religion and politics in Kenya. This will be preceded by a global description of a general-

Our approach can identify specific forms of speech associated with increased risk of violence, and allow organizations to indicate to observers where published speech is associated with increased risk of violence.

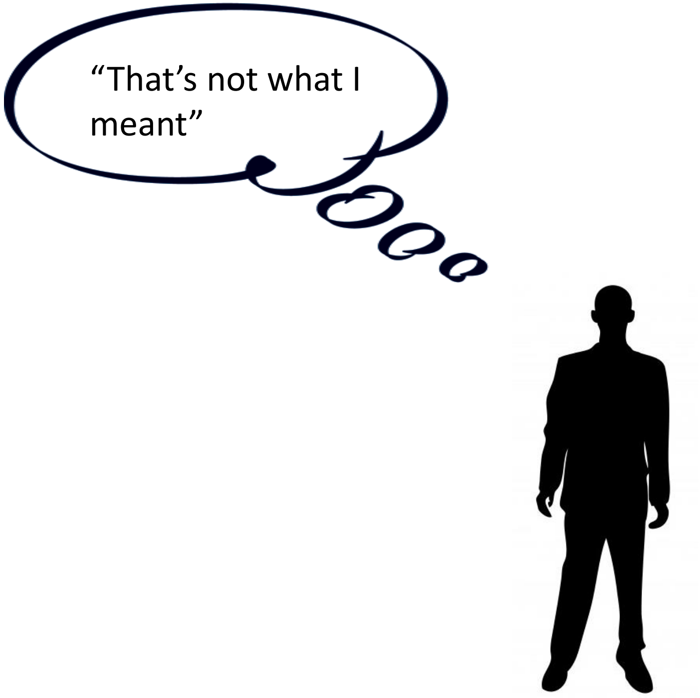
!

Instead of censoring the speech, publishers might indicate the level of risk associated with speech to audiences by publishing it alongside the speech.

WARNING: Language Associated with increased risk of violence

Furthermore, it is sensitive to assert that language incites people to violence because the intent of the language is subjective (in its deployment and interpretation).

Influential actors often assert different meaning to their language when confronted with its alleged incitement.



Nonetheless, the method will allow us to say that scientific evidence tells us that the language used is statistically associated with an increased or decreased risk of violence.

Previously, we have had only subjective interpretations of language and subjective interpretations of its link to violence.

Only a small portion of people understand how these technologies work, and an even smaller elite manages them. However, they affect and even disrupt the lives of many.

As enormous amounts of information increasingly concentrates in fewer and fewer places, we can expect calls to address issues of data control to increase.

This raises the importance of data use ‘explainability’ – capacity to identify *the precise language* associated with increased risk, not simply *the allegation* that someone’s language is associated with increased risk.