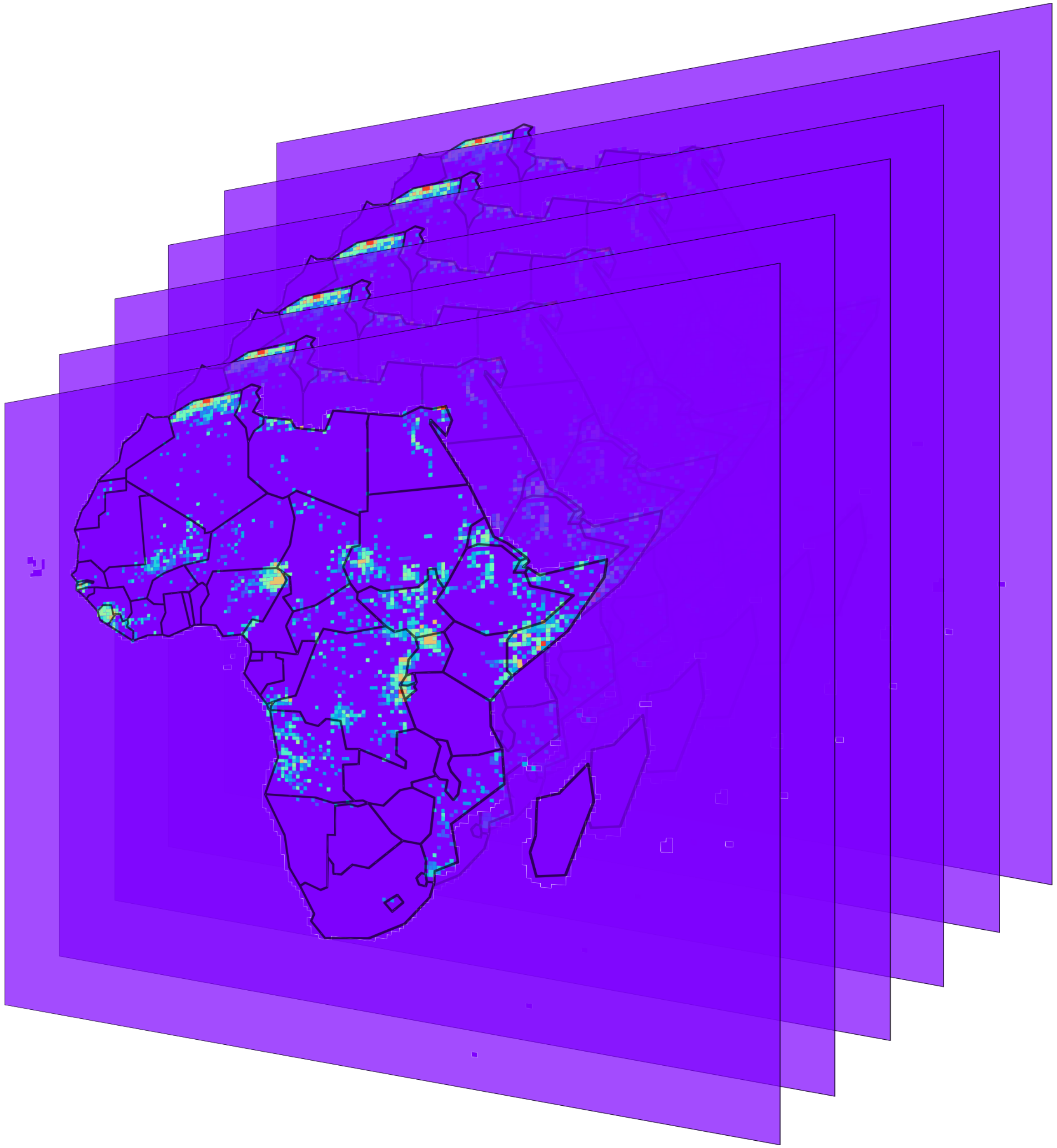
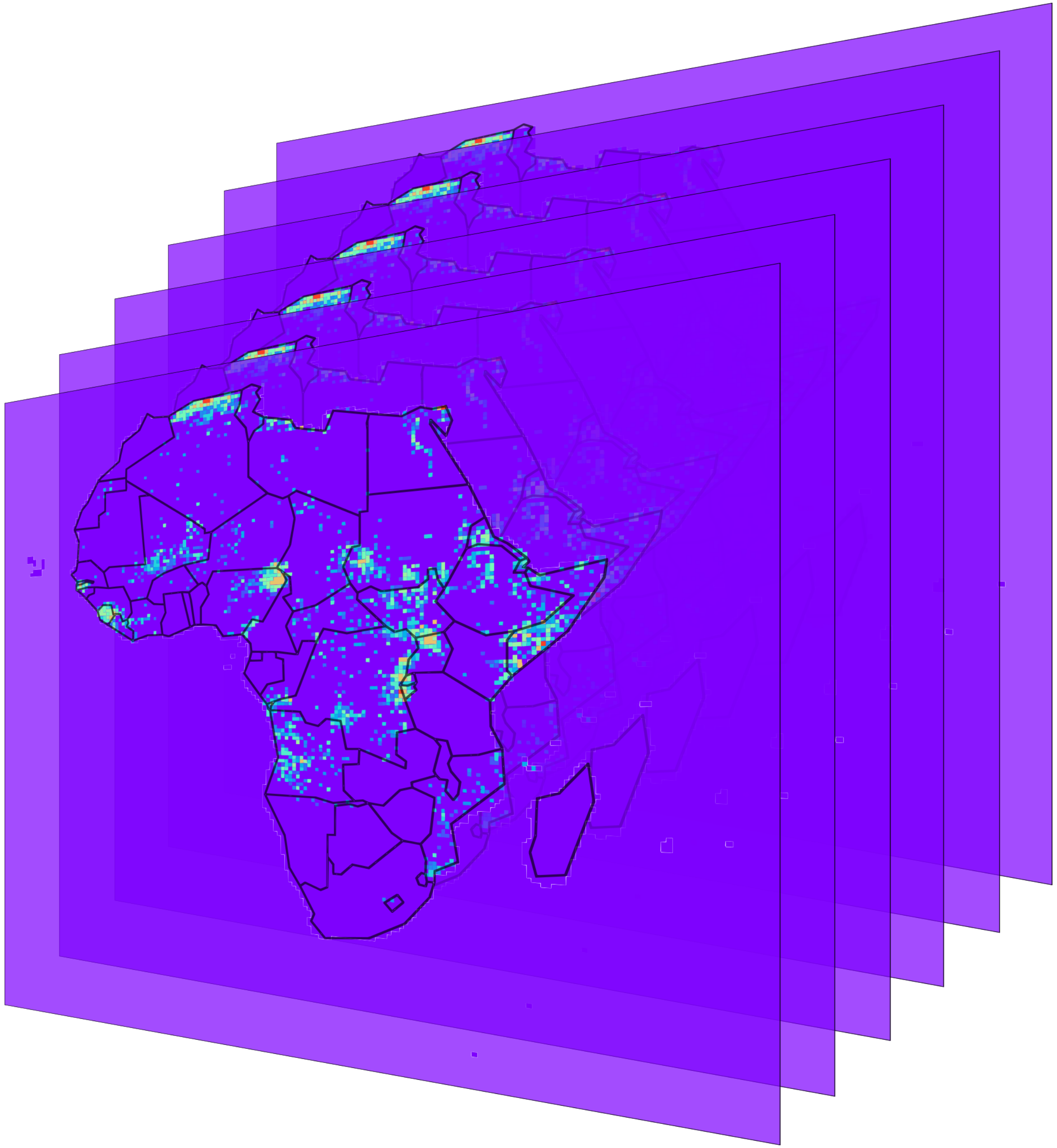
Anticipating Escalation

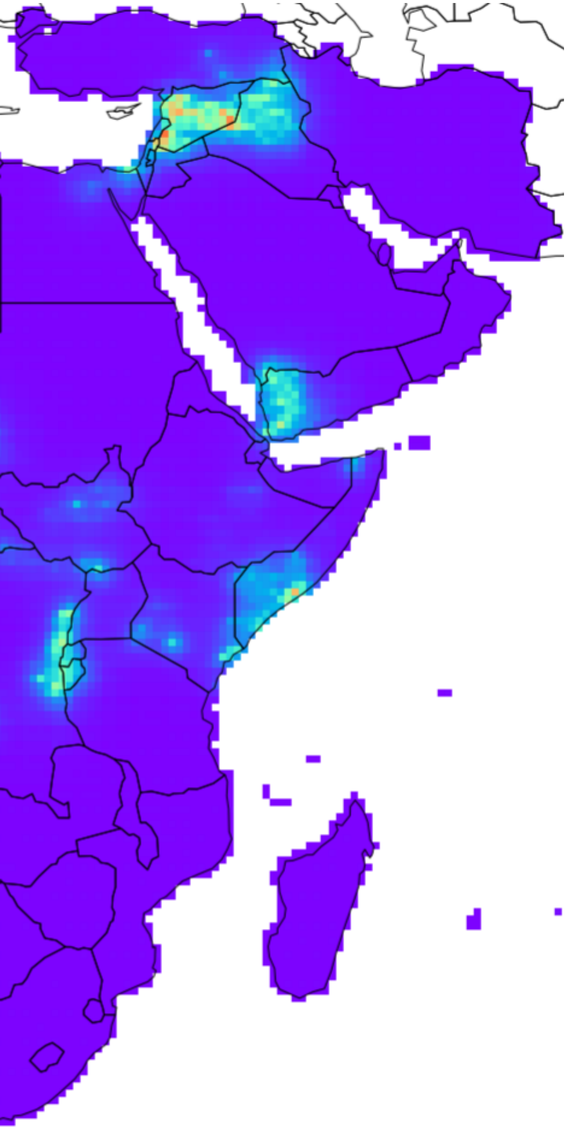
Actionable Insights with Actor-Embeddings and Transformers in Conflict Forecasting

w/ Mihai Croicu



Very preliminary - basically a pitch...



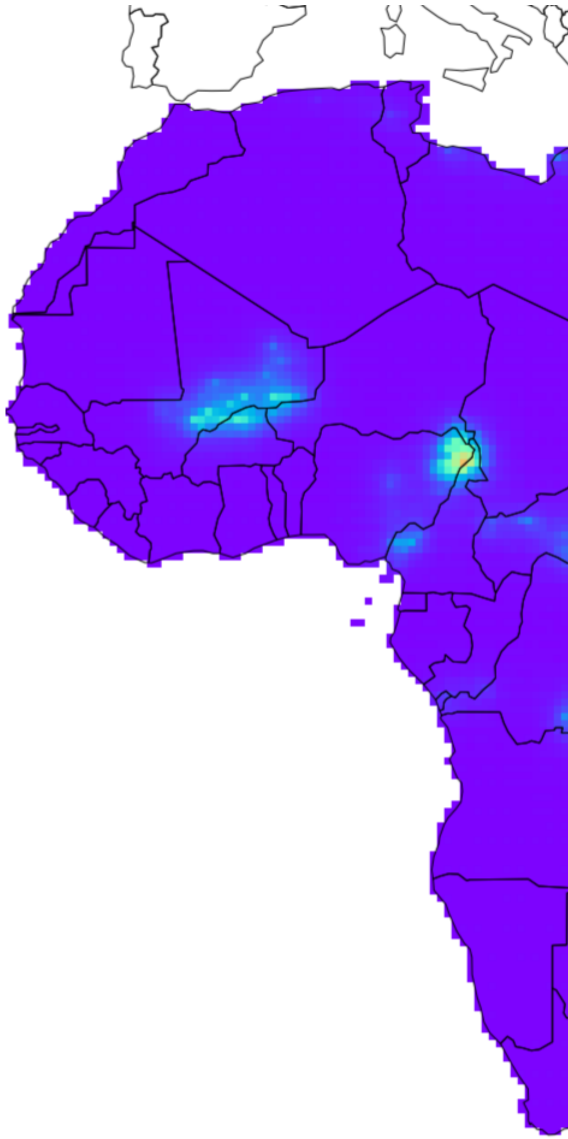


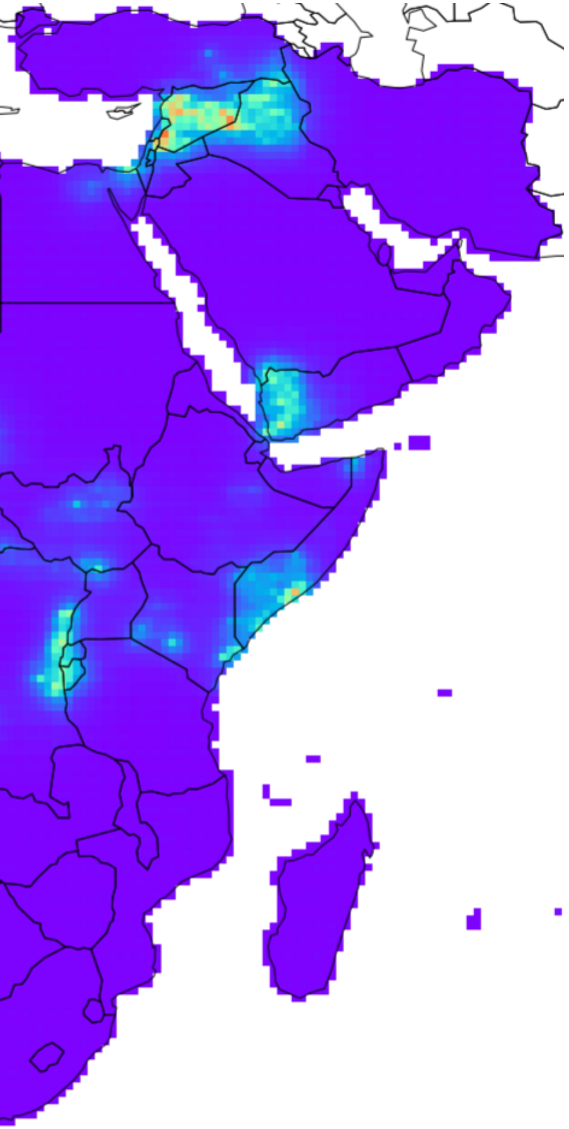
VIEWS already produces innovative and pioneering conflict forecasting

We have a robust **network of partners** who also work on conflict forecasting

Practitioners are experimenting with their own **early warning systems**

But most, if not all, current approaches have one **detrimental feature** in common...





Seth Caldwell,

Data Scientist at UN OCHA Centre for Humanitarian Data and a member of the scoring committee for the VIEWS prediction competitions

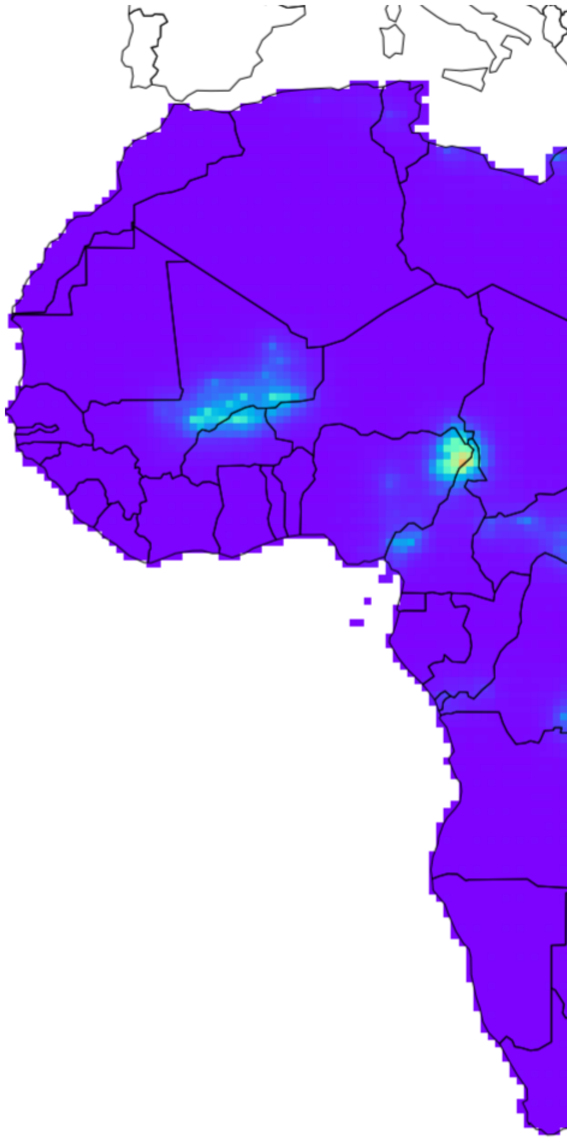
Has thoughts on this:

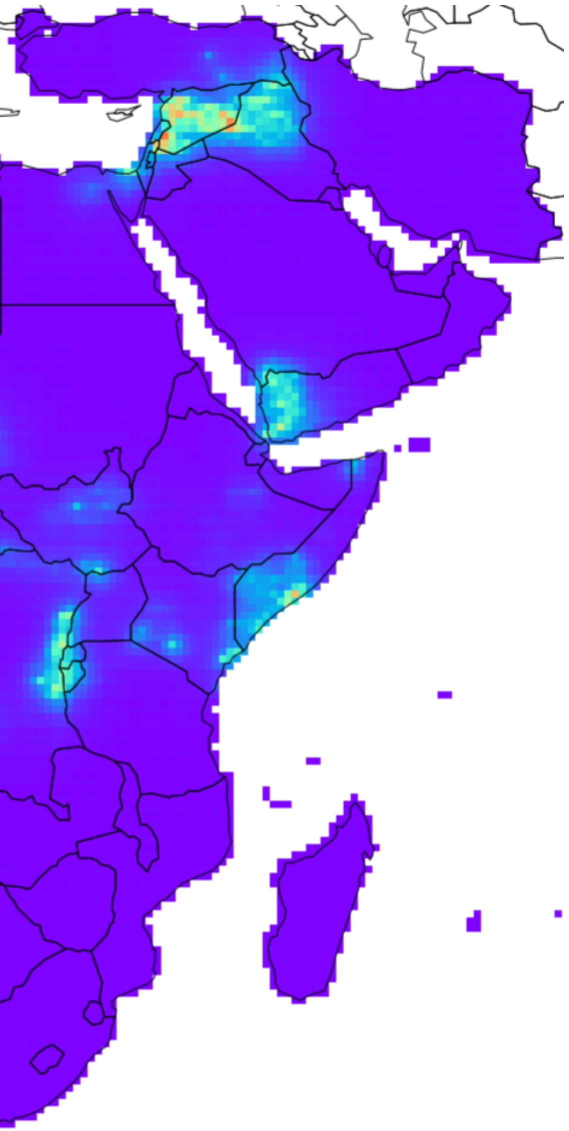
The thoughts:

> "[There is] insufficient justification for exclusively relying on conflict prediction models to drive anticipatory action due to several factors:"

- Poor performance in predicting the onset of new conflicts.
- The lack of clear connection between predicted conflict and resulting humanitarian impact.
- The dominance of ongoing conflict as a predictor of future conflict.

Source





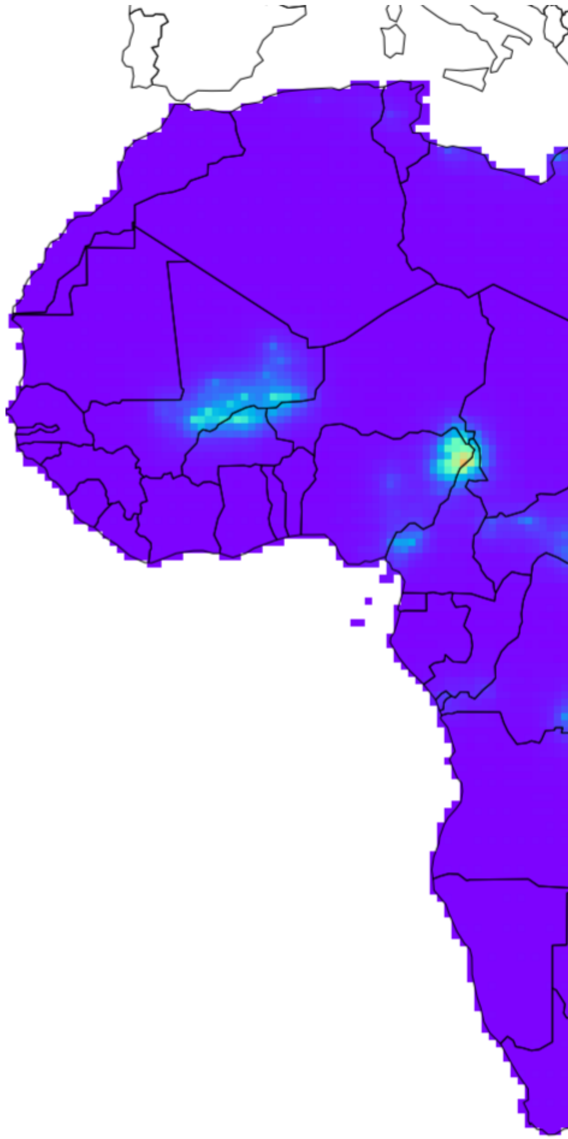
And indeed are **challenges**:

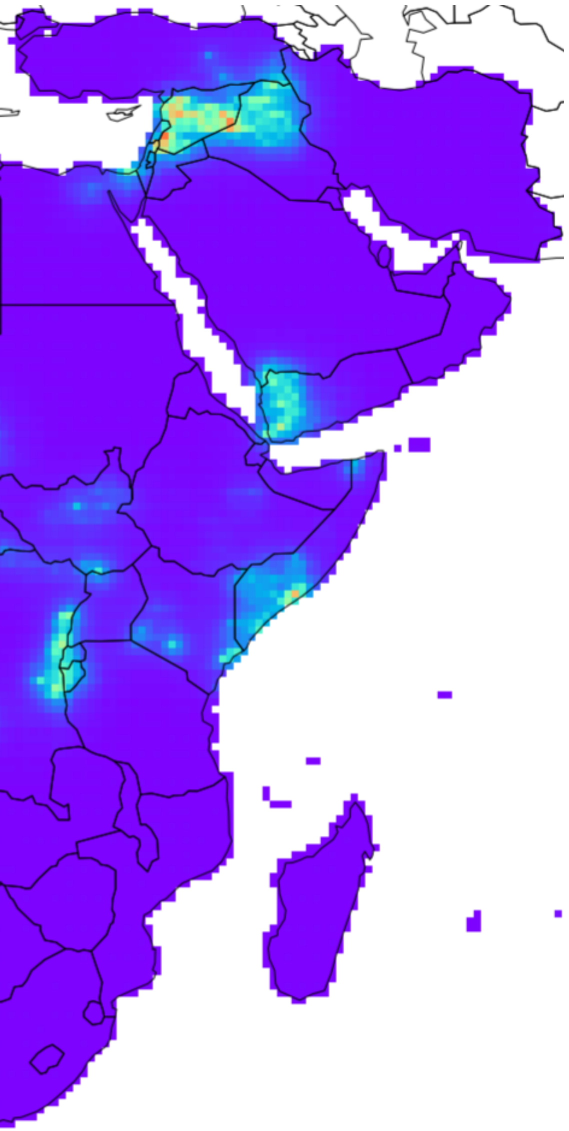
- **Conflict history**, in some form, is the primary predictor in most current systems.
- Socio-economic, institutional, demographic, and geographic features are usually aslo included in models but **lack varience in signals** due to inherent inertia or granularity of measurment.

Among a number of **recommendations**, Seth and his team rightly note that, for our models to be useful for anticipatory action in the humanitarian sector, we should:

> **"Focus models on predicting shifts in conflicts, such as an increase in intensity or onset."**

Source





Fair is fair, and we do hold Seth in high regard.

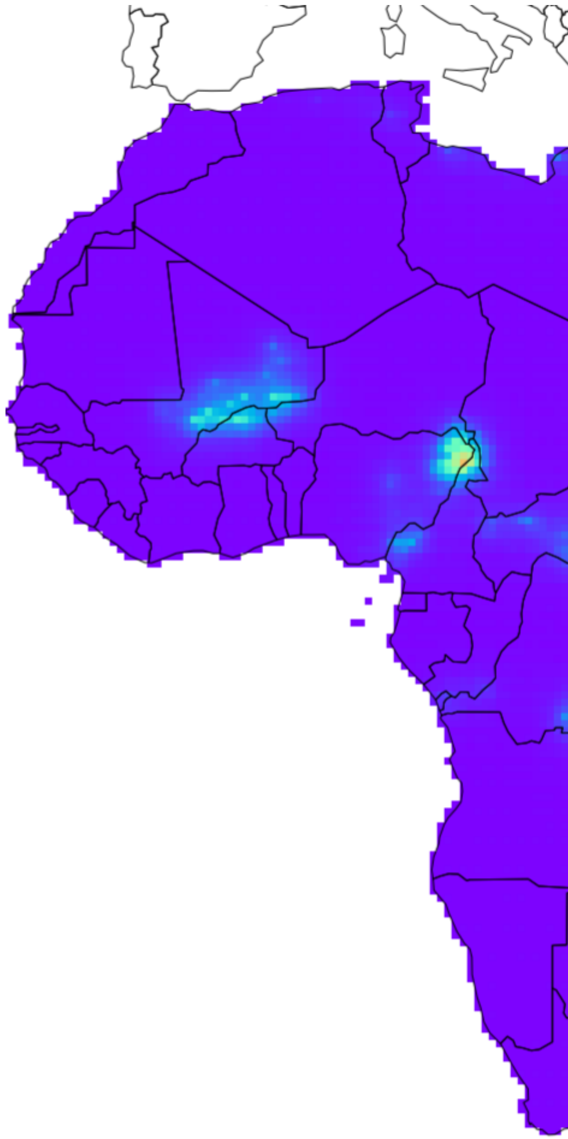
So now we are creating a **new model** for Seth.

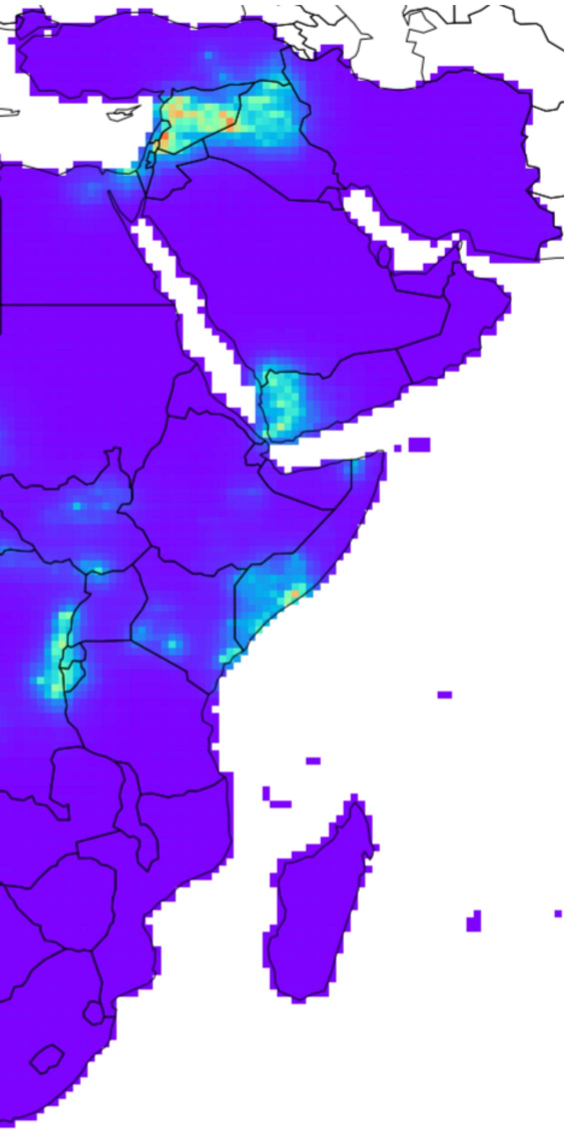
A model specifically designed to **generate actionable insights** by utilizing actor-embeddings and transformer networks to forecast **escalatory patterns** in violent conflict.

When desgining this new model, we will emphasize **escalatory patterns**.

The fact that we are **not exclusively focusing on onsets is deliberate**.

Forecasting onsets is considered the holy grail by some conflict scholars, **but...**





Mali, Chad, Guinea, Burkina Faso, Niger... **Limited** room for action.

Acting on anticipated escalations in conflict zones where humanitarians and international actors are **already present and mandated** is likely more realistic.

But **"both is good"**, so we will try to achieve both.

In a nutshell:

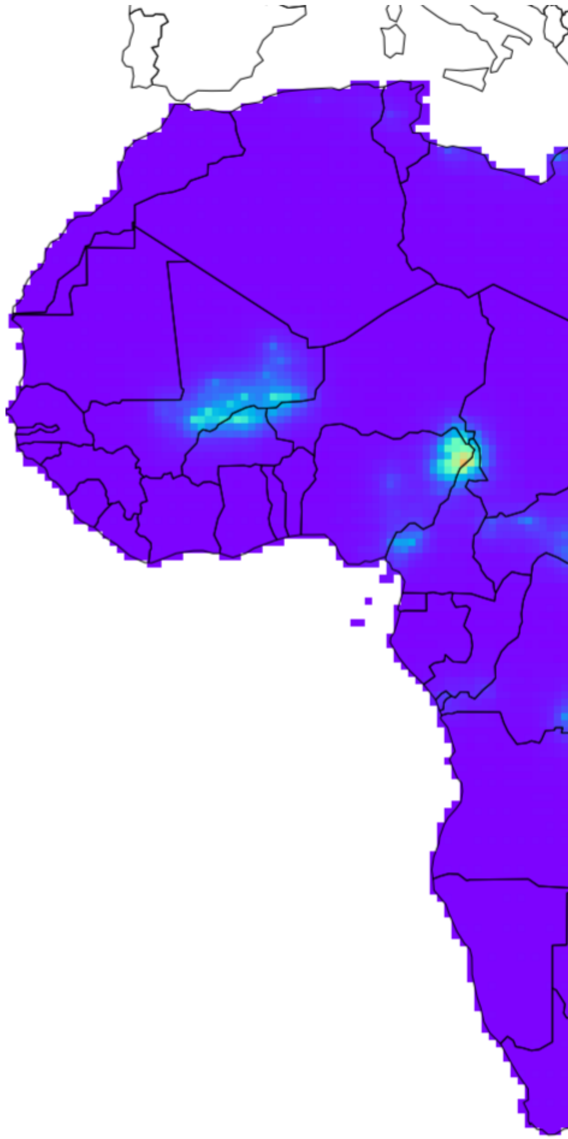
🔎 The unit of prediction will be **armed actors** (groups).

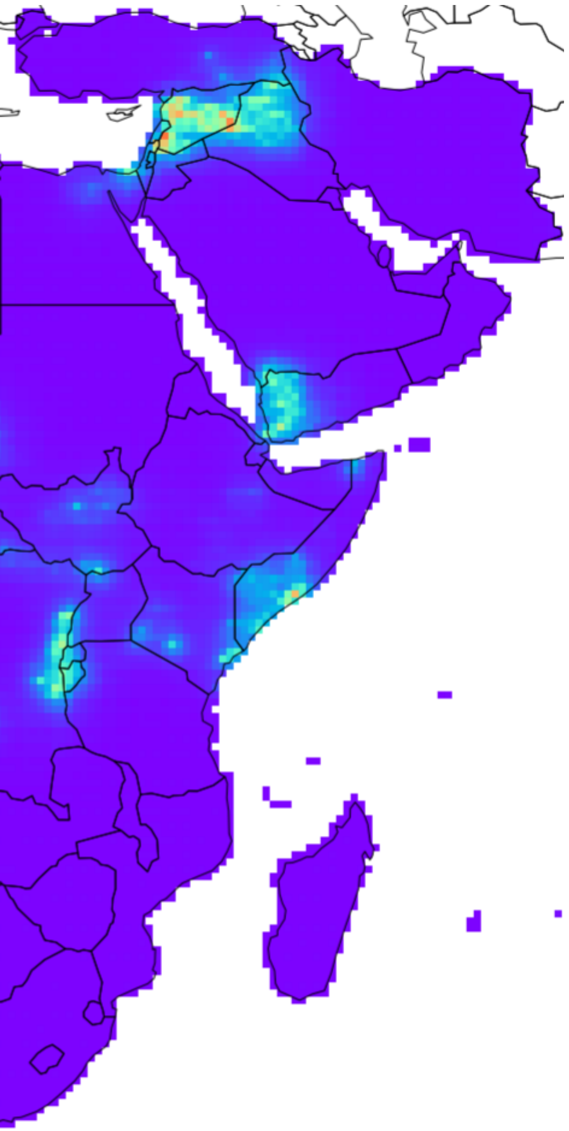
🎯 The prediction target will be the number of future battle-related **fatalities** produced by a given actor each month.

📈 We will forecast a **sequence** of 12 months.

🤖 To create these forecasts, we will use **transformer neural networks**.

📰 And as input for these transformers, we will create **actor embeddings** using text data (news sources, wikis, etc.).





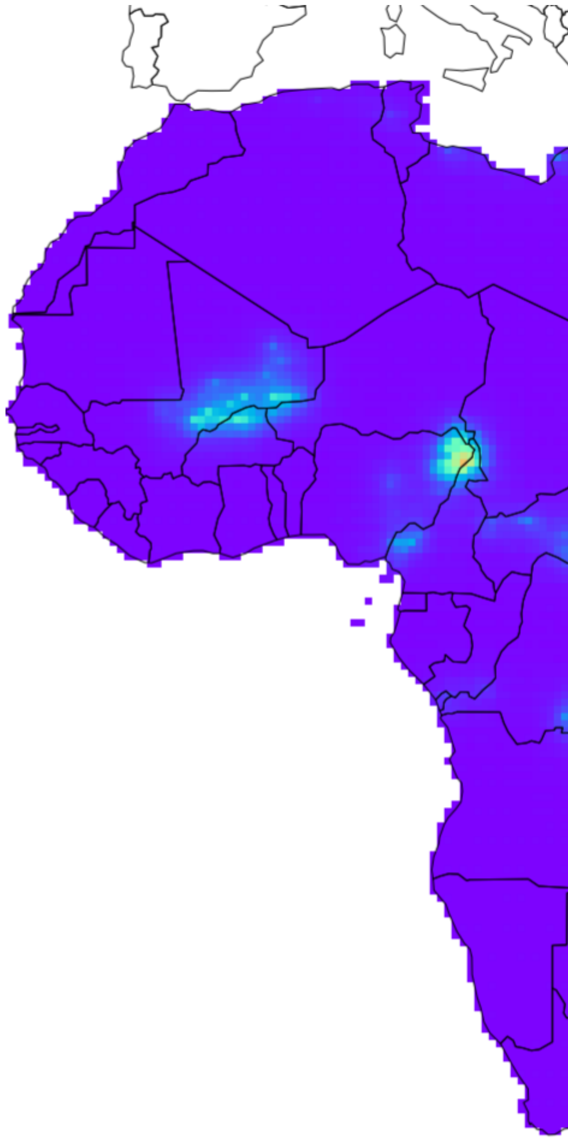
Lets unpack some of this jargon, specifcally

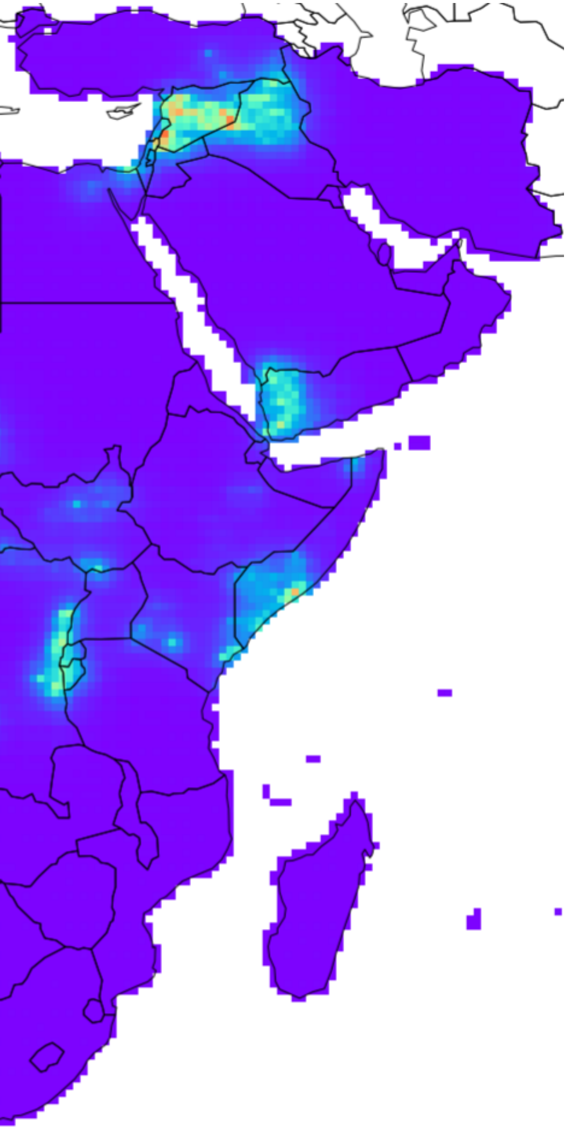
- **Embeddings**
- **Transformers**
- **Sequence**

Embeddings (1):

Embeddings act as **multidimensional summaries** of information. For instance, consider a 2D actor space with a capability and a ideology dimensions.

Now, imagine having an **actor space** with 1000 dimensions for a much more nuanced summary.





Embeddings (2):

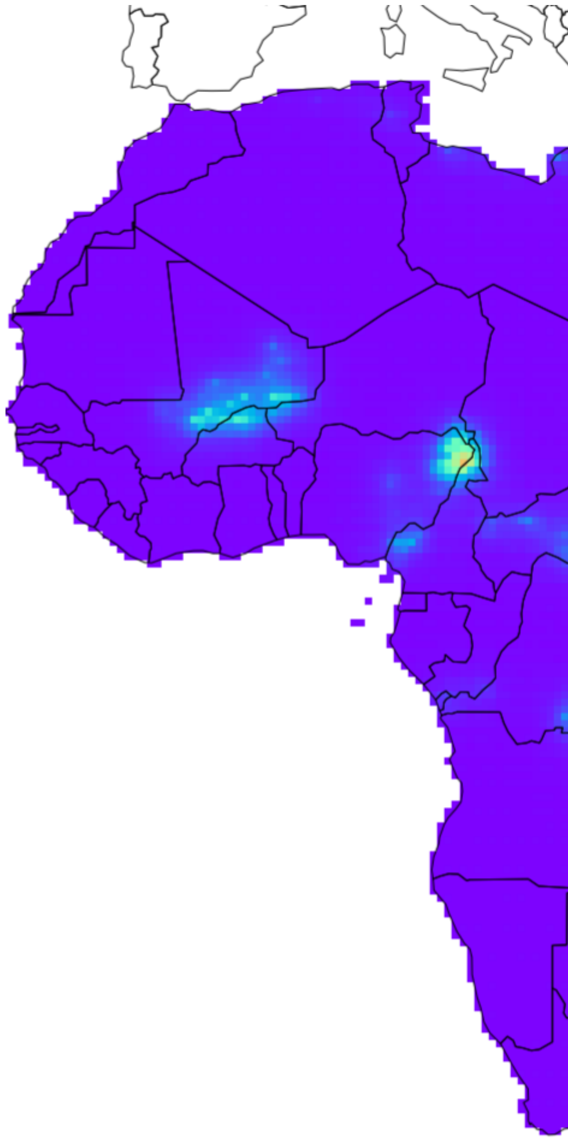
Machine learning models are trained using large amount of **text data** to position armed groups (relative to each other) in a highly multidimensional space.

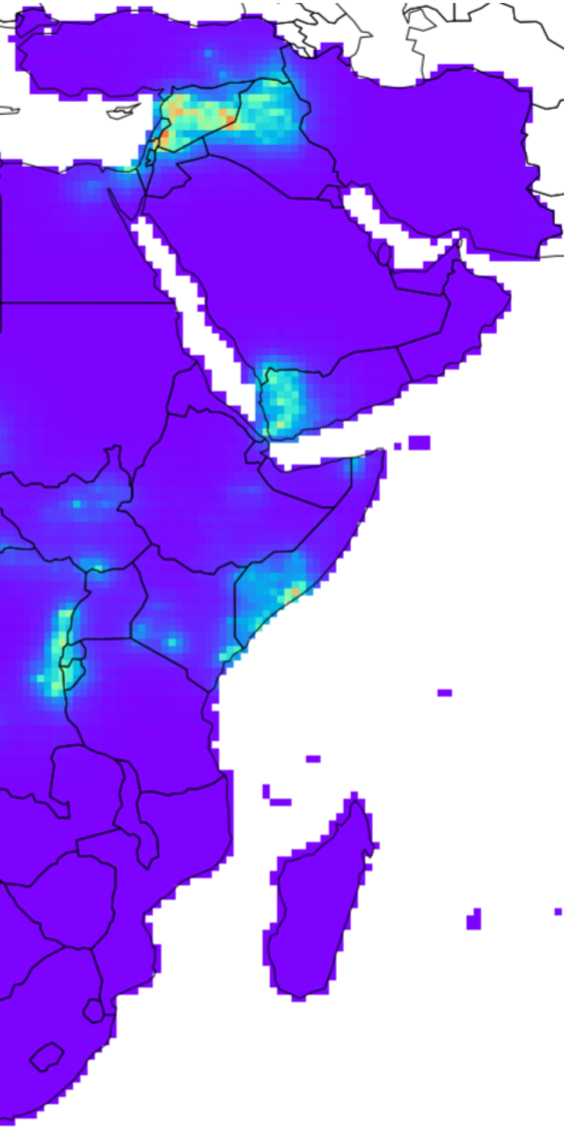
Now, Actors are no longer just names or numbers. They are represented by vectors in this multidimensional space - but their **positions can change over time**.

Transformers/Attention (1)

Transformers take the multidimensional **actor embeddings as inputs**.

Transformers employ an attention mechanism to learn **relationships and similarities** between the input embeddings - which here represents different actors at different times.





Transformers/Attention (2)

When "asked" to review a given actor, the transformer compares it to itself and to **all other observed actors** across time.

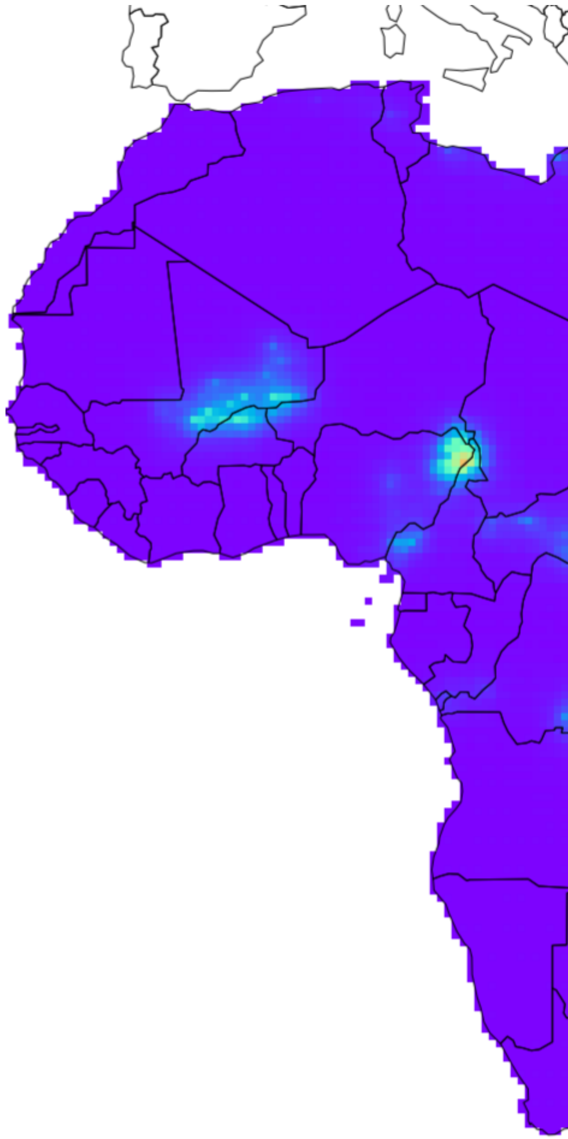
By discerning these patterns, the model constructs the **most likely future pattern for the actor** under review, (hopefully) enabling the model to anticipate violent actions.

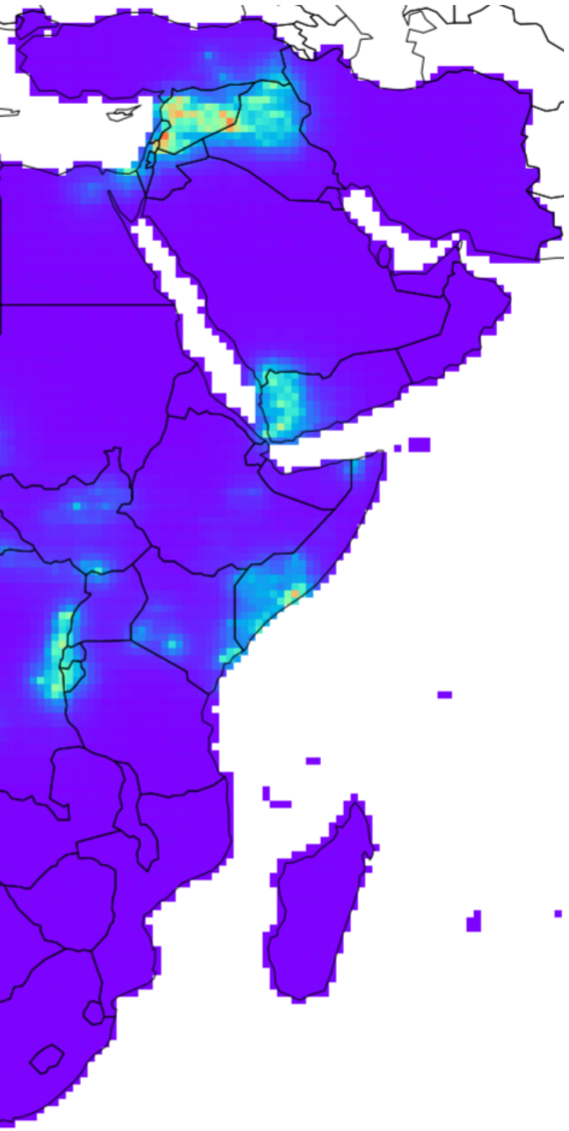
Sequence

Notably, the target and thus the output will be **a sequence of months**.

Current research suggests that the **stepshifter** models and **autoregressive** models struggle to discern between temporal units.

I.e., the probability mass is spread across time. **Forecasting a full sequence** at each step might alleviate this."





Additional features (1)

Beyond the goal of achieving accurate predictions, this approach helps in incorporating new actors into prediction frameworks.

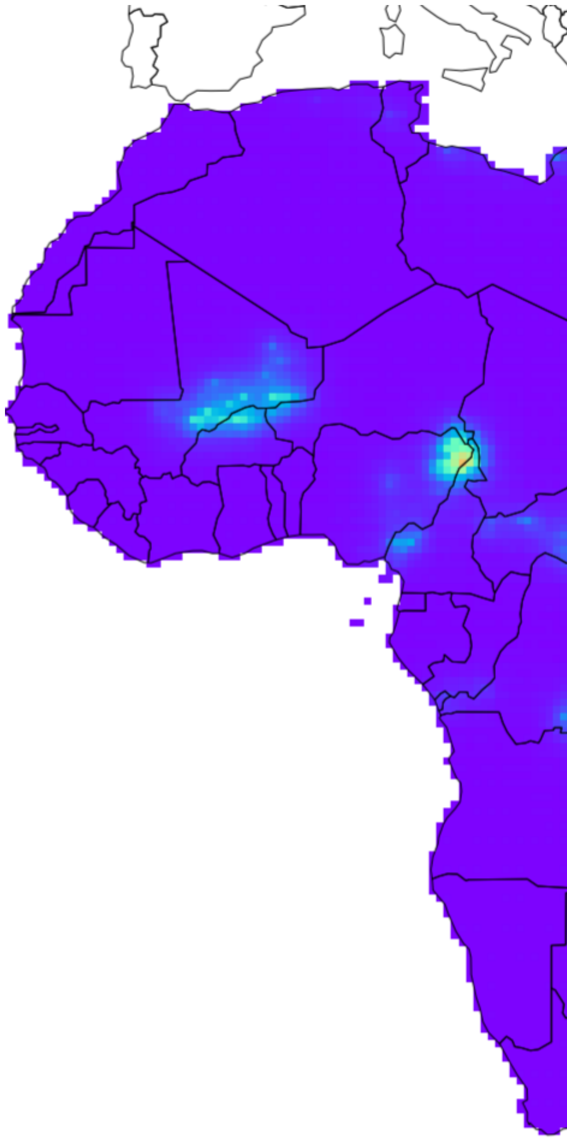
I.e, it is challenging to assess whether a new actor will exhibit behaviors similar to one or more known actors and indeed to which known actors.

Positioning new actors within an actor space, which can be continually updated, serves as a valuable (hierarchical-ish) prior.

Additional features (2)

Furthermore, examining the relationships among established actors in embedding space might be a worthwhile study in itself.

I.e. understanding how these known actors relate to each other in this multidimensional space is likely of substantial interest.



Mihai Croicu &
Simon Polichinel von der Maase

📫 simmaa@prio.org
🐙 https://github.com/Polichinel
🏰 PRIO, Oslo