



# IITSEC 2024



## Evaluation of Historical Journalism Data for Decision-Support System Gray-Zone Models

# CAE

Colby “Blade” McAlexander / CAE USA

Jay Freeman / CAE USA

# Table of Contents

- Lexis
- Introduction
- Terms of Reference – Paper
- Terms of Reference – Project Framework
- Data Pipeline and Offline Training System
- New Time Dimension
- Data Normalization and Correlation
- What Else Can this Data Provide – (No Plan Meets First Contact)
- Predictive Power – (For Models or Unfolding Events)
- What Else Can this Data Provide

Minimizing acronyms; however, a few acronyms will be used for brevity and understanding.

## Data Store for News Sources

- **GDELT** ..... Global Database of Events, Language, and Tone

## Instruments of National Power

- **DIME** ..... Diplomatic, Information, Military, Economic
- **DIME-FIL** ..... DIME + Finance, Intelligence, Law Enforcement

## Analysis of an operational environment, the threats and opportunities within it, and how to respond

- **PMESII-PT** ..... Political, Military, Economic, Social, Information, Infrastructure, Physical Environment, Time

# Lexis

## Gray-zone Lexicon

## Behaviors/ Actions

- 👤 **Potentially leading to, but below the threshold of armed conflict**
- 👤 **Contrary to widely accepted/ratified responsible norms/behaviors**
- 👤 **Actors have knowledge on how opponents perceive its actions**

Employed across competition and armed conflict continuums, with/with-out employing the military



## Gray-zone Activities

## High-Level Examples

**Common Across DIME-FIL, Cyber, Space:** espionage | influence operations | disruption | introduce ambiguity/bias

- 👤 **Diplomatic**....election meddling | malign influence
- 👤 **Information**...news & social media | data breaches
- 👤 **Military**.....border incursion | disputed territory
- 👤 **Economic**....foreign investment coercion | debt-trap

- 👤 **Financial**..... foreign business pressure | bribery
- 👤 **Intelligence**..... espionage | coercion
- 👤 **Law Enforcement**...foreign secret police | detention

- 👤 **Cyber**.....communications provision ⇔ control | social media apps | intellectual property stealing | supply-chain
- 👤 **Space**.....co-orbital maneuvers | anti-satellite weapons 'testing' (space and ground-based) | jamming

# Introduction

## Joint Concept for Competing 2023

- "Analyzing any adversary's way of war is instructive."
- "We think of being at peace or war... **our adversaries don't think that way.**" — General Joseph F. Dunford
- "They believe they are in a **long-term "conflict without combat"** to alter the current international system..."
- "Strategic competition is thus an **enduring condition to be managed**, not a problem to be solved."

## Long-term Conflict w/o Combat Strategy



## Alter the System

- **Establish new** norms, standards, behaviors, 'red-lines/-zones,' 'rules of the game'
- Methodical, slow **whittling of opponents' center(s) of gravity** ⇒ **'court of public-opinion' | imposing costs**

## Modeling & Simulation



### Explore gray-zone tactics in terms of strategic 'soft-power' campaigns & effects

- DIME **conceptual modeling** to inform decisions & consequences of actions
- Build a **process to construct topic models** to characterize DIME actions



# Terms of Reference – Paper

## Purpose



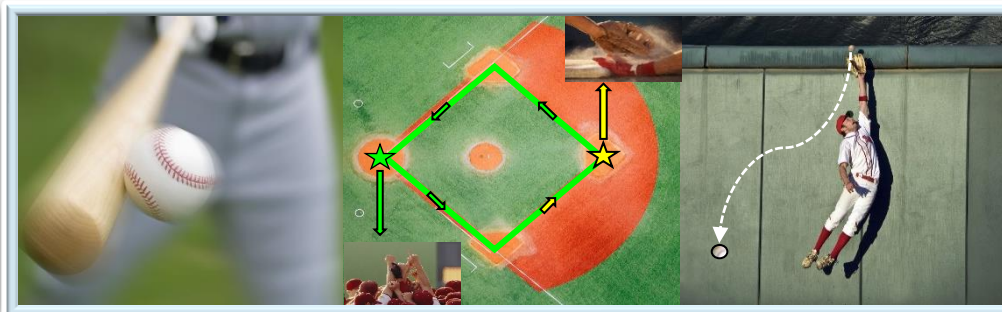
Examine the results of feeding GDELT newsfeeds into 5-dimensional statistical DIME space + time as a basis for risk-cost-benefit calculations in the gray-zone arena.



## Guiding Questions

- **New time dimension (0.25 to 8.25 years)**
  - ❑ Has data changed over time?
  - ❑ What are the driving forces?
- **Data correlation and normalization**
  - ❑ Does it help identify a topic of 'interest?'
  - ❑ Or a theme of 'interest?'
- **Predictive Power (for models or unfolding events)**
  - ❑ Determine behavior signatures and likelihood?
  - ❑ To what level: operations; training; wargaming; \_\_\_?
- **What else can this data provide?**
  - ❑ Journalism as a sensor?

## No Plan (or Research) Meets First Contact



- In research always desire to hit the home run on first pitch ...;
- **However, by definition, research—a systematic investigation—can force running the bases.**

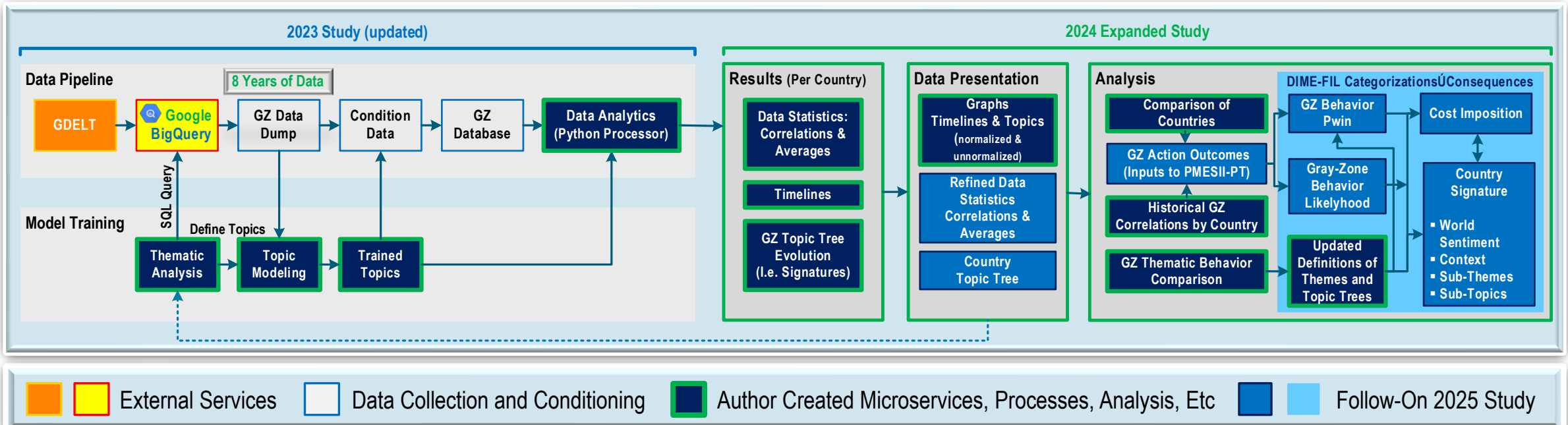
# Data Pipeline and Offline Training System

## Updated and Projected

Saz-Carranza et al (2018) find based on timespan, coverage, and translation, “along with its constant content update and its gradual inclusion of non-Western newspapers, have earned GDELT a positive degree of reliability among scholars.”

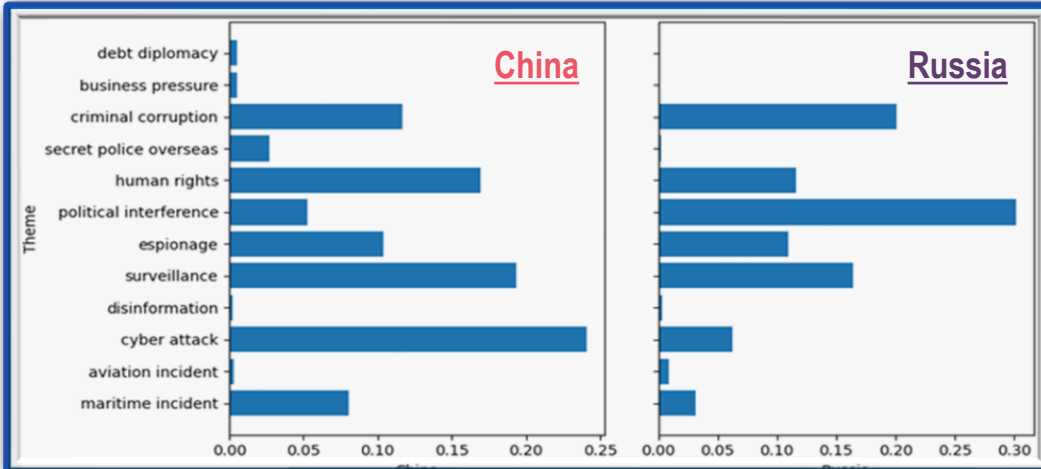


- **GDELT’s quantitative analysis potential attributes to it’s big data**
- **Employ best practices from social sciences (e.g. psychology)**
  - ❑ Thematic analysis done haphazardly or without ‘checksums’ could further mask ground-truth
  - ❑ Creating something which does not exist ⇔ ⇐ Missing which does exist

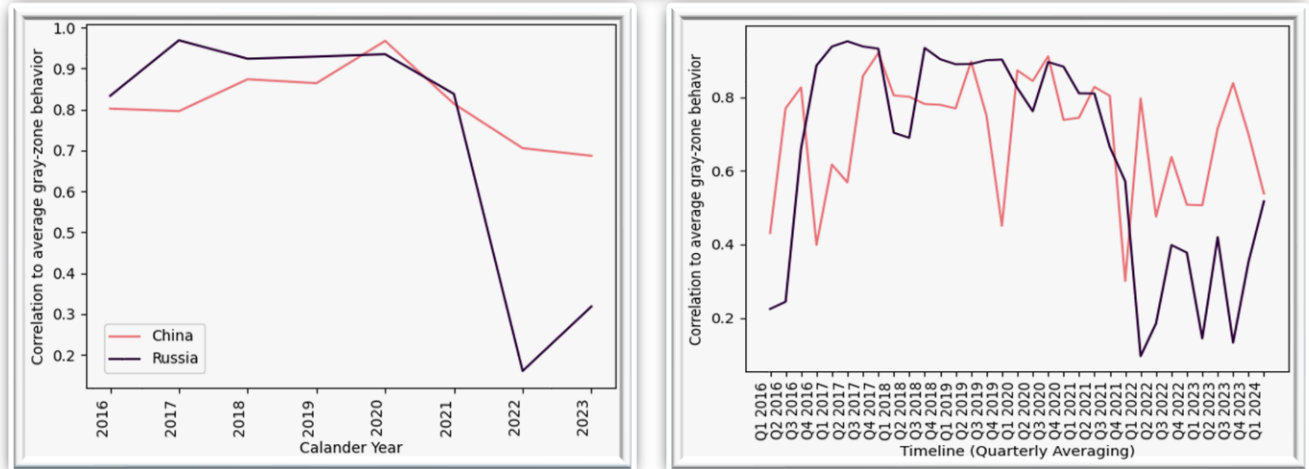


# New Time Dimension & Data Normalization and Correlation

## Normalized Thematic Signatures Averaged Over 8+ Years



## Correlation between Annual and Quarterly Gray-zone Topics to 8-year Average



## Topics initially created initially by each month

- Add mentions like vectors to create country's behavior representation
- Applicable over more extended periods of time

## Pearson Correlation

- Gives precise measure of linear correlation

## 1<sup>st</sup> paper (0.25 years data pull) could not definitively answer

- Are countries' gray-zone thematic signatures consistent from year-to-year, or do they evolve over time?

## This paper (8.25 years data pull)

- Yearly averages correlate well with thematic signature; quarterly are noisier
- China: consistency; annualized signature well correlated to baseline
- Russia: consistency from 2016 to 2021; 2022 very noticeable phase shift

# What Else Can this Data Provide – (No Plan Meets First Contact)



**Data Pull = 55 TB (45+ years) down to 30 GB (8.25 years)**

- Size more manageable, but still large
- Still very unstructured data

**Sifting Through Data 'Chaos' for 'Actionable' Information**

## A Makeshift 'Portal' Required

- To facilitate drill down analysis
  - ❑ Better understand timeline graphs
  - ❑ Automate deeper lower-level queries
- 15 min increments (GDELT update rate)

## Source URL may NOT open a web page

- Not all sources web-based
- Some web-based sources removed from site

```

1 {
2   "general": {
3     "SOURCEURL": "http://donpress.com/news/20-03-2016-rossiya-gotovit-zhitelyan-okkupirovannyh-territoriy-kiberterror", "V2Themes": "MANMADE_DISASTER_IMPLIE
4     "SOURCEURL": "http://runun.es/opinion/246376/informe-otalvora-guyana-ejecuta-agresiva-ocupacion-de-territorio-venezolano.html", "V2Themes": "MANMADE_DI
5     "SOURCEURL": "http://nation.com.pk/international/25-Mar-2016/china-launches-charm-offensive-for-first-overseas-
6     "SOURCEURL": "https://russian.rt.com/intv/intv/2016-01-31/Le-Monde-Rossiya-usilivaet-kontrol", "V2Themes": "IMMIGRATION,1425;TAX_FNACT_IMMIGRANTS,
7     "SOURCEURL": "http://freerepublic.com/focus/f-news/3394826/posts", "V2Themes": "EPU_CATS_MIGRATION_FEAR_FEAR,971;EPU_CATS_NATIONAL_SECURITY,67;EPU_CA
    }
  }

```

138,256 URLs

~55s URL ⇒ | China | Behavior = Surveillance | Outcome = Military | Mentions

Quarter 1, 2016

The screenshot shows a web browser window with a Reuters article. The article title is "China launches charm offensive for first overseas naval base" by Ben Blanchard, dated March 24, 2016. The article text mentions that China has launched an unusual charm offensive to explain its first overseas naval base in Djibouti. A photo shows a soldier in a blue uniform standing on the deck of a ship. The browser's address bar shows the URL: http://www.reuters.com/world/china-launches-charm-offensive-for-first-overseas-naval-base/.



# Predictive Power (for models or unfolding events)

## Influencing Legitimacy and/or Pushing Boundaries

- Mentions can train statistical models, capturing behaviors in noteworthy & actionable context
- Each journalist report is > a vote casted for the event's newsworthiness
  - ❑ Provide signatures and periodicity of similar actions with context
  - ❑ Potential measure of legitimacy in terms of an actor's actions
  - ❑ Creates DIME-FIL links between behaviors and consequences
  - ❑ Local awareness—may have missed due to 'butter-fly effect' or swift-moving event

### Analogy to 'Predict' the Response of 'Markets' to Patterns World State Data

#### Hedge Fund Companies

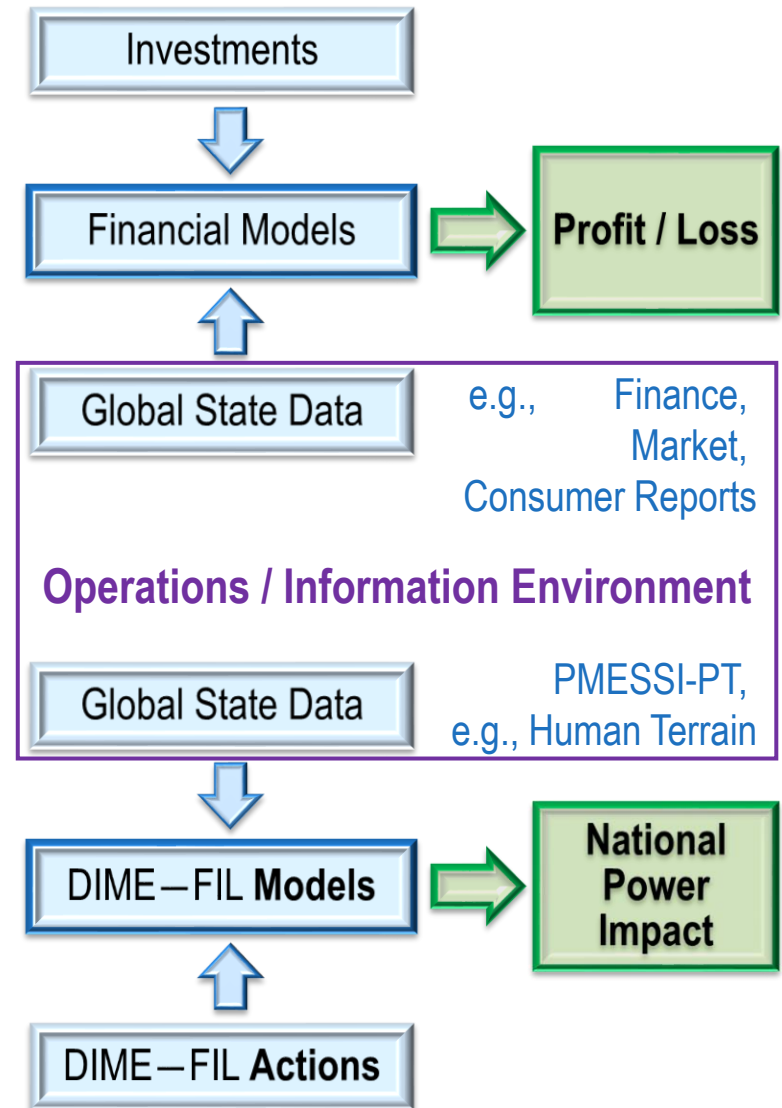
- Response to market
- Determine if a particular investment leads to profit or loss for the investors

#### Nations States / Global Actors

- Response to influence and legitimacy
- Determine if an instrument(s) of national power affect overall world standing

Fundamentally competing for finite resources and 'trading commodities'

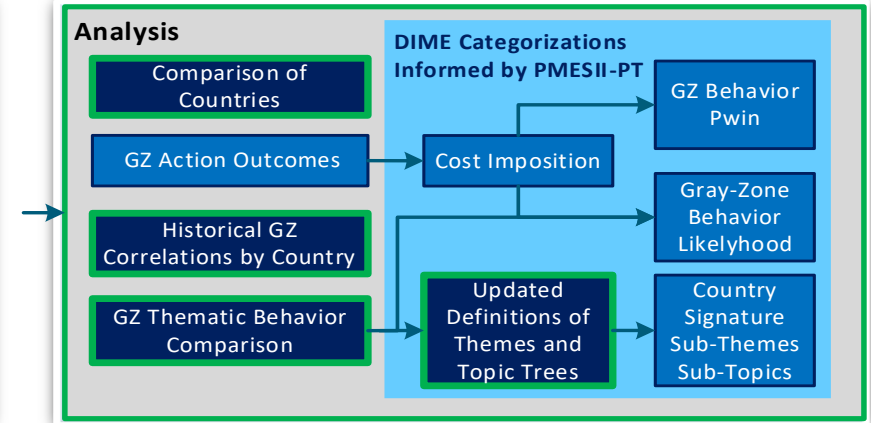
How their use of currency will affect their overall standing in the market / world



# What Else Can this Data Provide

## While journalistic data, alone, can't wholly inform trade-space analysis, it can

- Provide context, tones, and sentiments
- Assist in distinguishing pattern anomalies
- Provide meaningful understandings of human behavior and beliefs
- Capture the above in terms of global relation interdependencies



## Operations

- Insatiable appetite for data can overwhelm
- Speed of Fight? ⇒ New analysis still needs exploration
- Atypical breaking events as they unfold ⇒ Yes !?!

## Training, Exercise, Wargames

- Shortens research time ⇒ pre-, during, post- event
- Provides normalized information ⇒ plausibility
- Break group think of over recycling



# IITSEC 2024



NTSA

## Questions





# IITSEC 2024



NTSA

## BACKUP Slides



# Terms of Reference – Project

## Situation



- Increasingly volatile, uncertain, and complex political-military strategic environments
- Adversaries employ gray-zone tactics to hold our national resources, instruments of power (i.e. DIME), and emerging technologies at risk

## Problem



### Today's decision analysis and planning solutions lack algorithms which

- apply search and statistical techniques to
- **make full use of journalism data and social media contrasted with**
- historical context, emerging media, and arising global societal-scale behavior

**to inform strategic decision-making.**

## Purpose

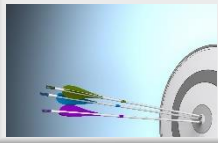


### Investigate news and social media sources to

- provide insight for creating frameworks and models for
- a future solution which feeds decision-support systems to
- aid decision-makers' examination of gray-zone tactics

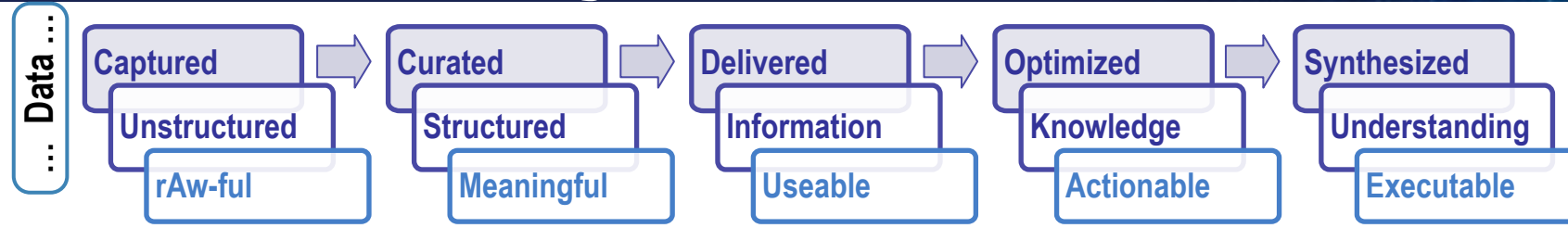
**in terms of strategic campaigns and effects**

## Objectives

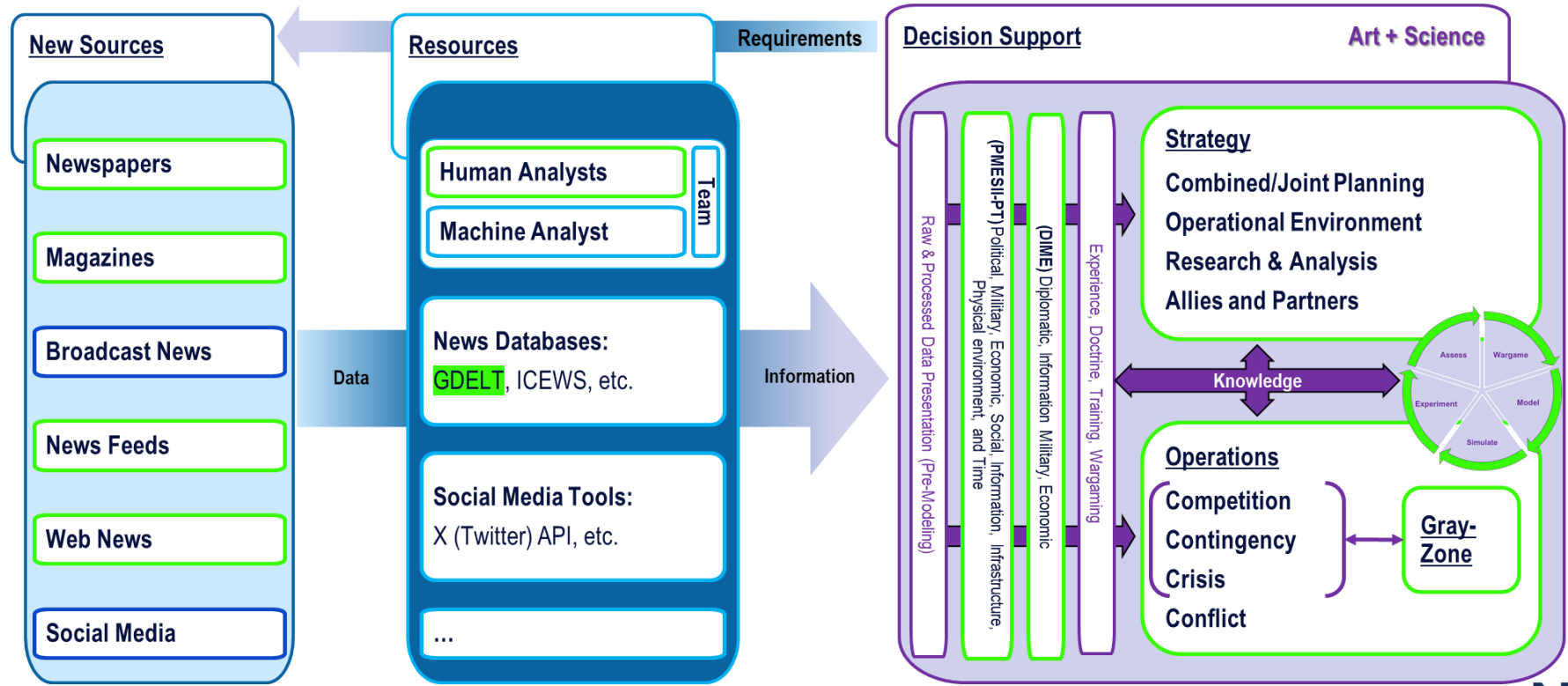


- ✦ Exploit journalism data to quickly organize ideas, explore issues, identify unknowns and analyze implications across multiple-domains, -actors, -geographies, -languages, -cultures, and -human sentiment.
- ✦ Investigate what news sources are actionable for use in 'soft-power' conceptual models (e.g., DIME; PMESII-PT).

# Terms of Reference – Project Framework



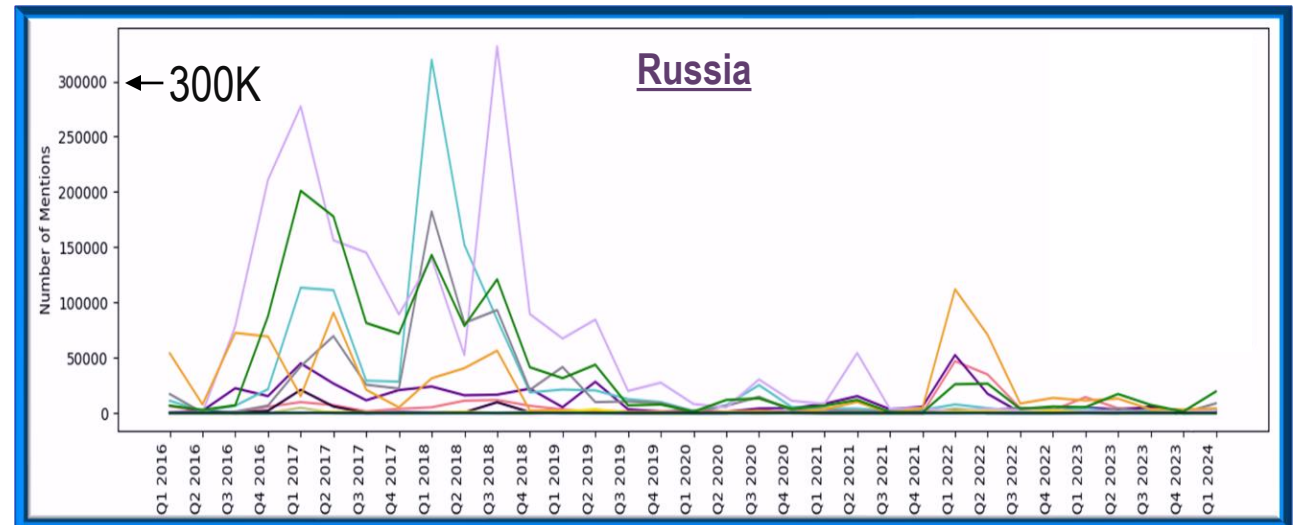
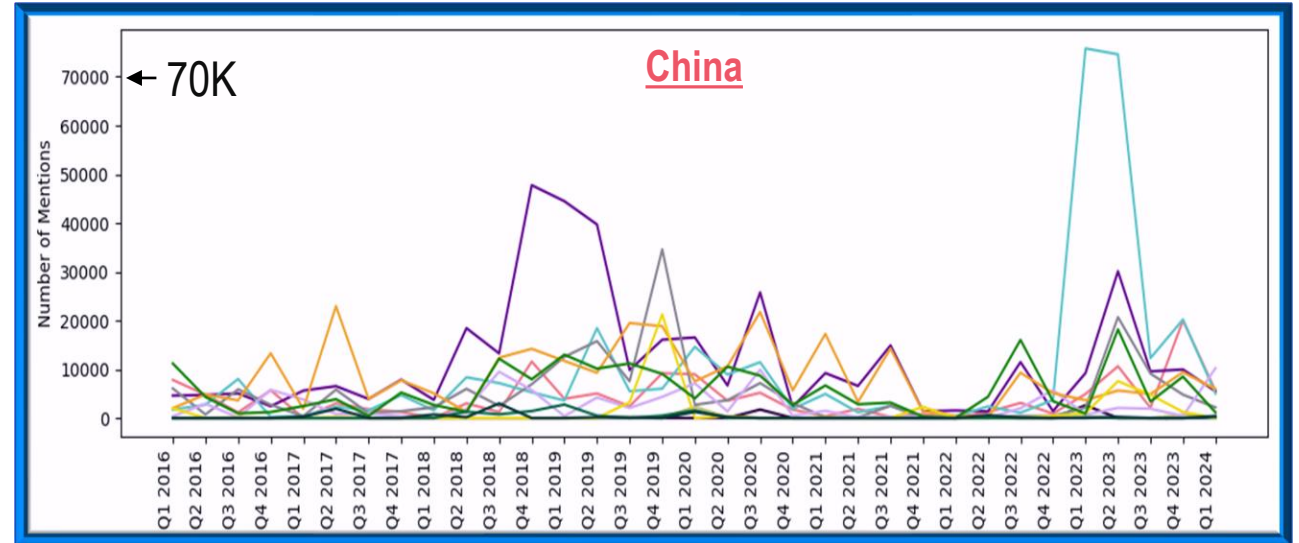
The struggle between tacit knowledge (art) AND confidence weighting (science) AND the desire for perfection



# New Time Dimension Behavior Breakdown

## Data from January 2016 to April 2024, depicted quarterly

- Number of mentions to news events associated with specific gray-zone themes
- Press coverage for Russia is 4x's higher than for China
- Top-2 Gray-zone Behaviors
  - ❑ **China**: cyber-attack and surveillance
  - ❑ **Russia**: political interference and criminal corruption



### One Event Stands Out, Towering Over All Others

- Russian interference in the 2016 US presidential election
- Runs for a three-year period

- maritime incident
- aviation incident
- cyber attack
- disinformation
- surveillance
- espionage
- political interference
- human rights
- secret police overseas
- criminal corruption
- business pressure
- debt diplomacy

