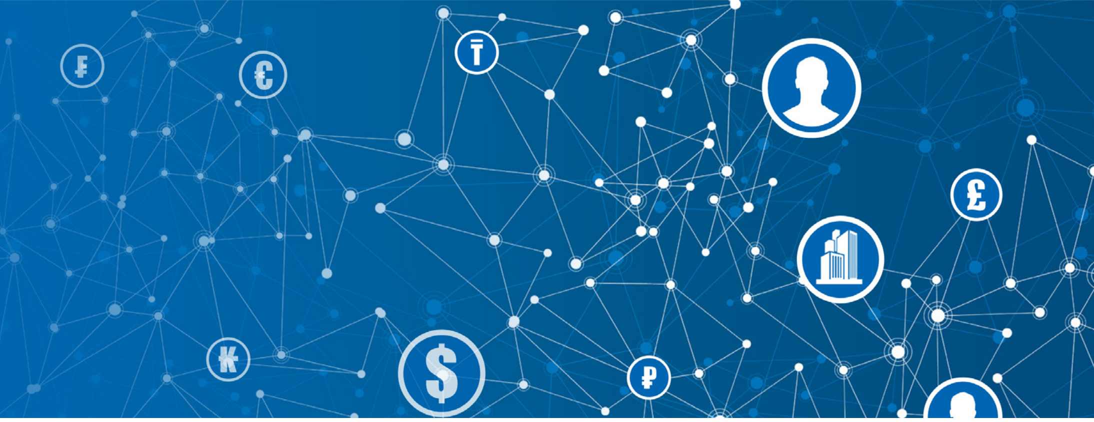
GOST

POWERED BY GIANT OAK



GOST ® For Elections Security (GOST-ES)

Background

Social media is being used by well-resourced foreign actors to spread political disinformation globally. Since 2014, there have been at least 60 distinct campaigns targeting more than 30 different countries. Over 70 percent of these campaigns have been conducted by Russia, with the United States as a frequent target. On April 20, 2020, the United States Senate Select Committee on Intelligence released its *Report on Russian Active Measures Campaigns and Interference in the 2016 U.S. Election*, in which it demonstrates a coherent basis to conclude that Russia interfered.

As the 2020 U.S. elections approach, it is imperative that all tools are used to combat political disinformation and ensure election security. As of March 2020, Giant Oak is tracking the 15 state-sponsored COVID-19-related disinformation narratives, with nine attributed to Russian campaigns. Because of their industrialized nature, state influence campaigns leave distinct signatures in social media content. Protecting elections while defending civil liberties requires two capabilities:

1. Rapidly understanding the scope, nature, and content of manipulation campaigns, and
2. Ensuring that efforts to address that activity do not stifle genuine debate.

GOST® for Elections Security (GOST-ES) accomplishes both requirements by incorporating new approaches to detecting coordinated online influence campaigns, developed by Giant Oak Distinguished Scientist Dr. Jacob Shapiro.

Solution

GOST-ES detects and characterizes political influence campaigns on social media using platform-agnostic machine-learning (ML) tools informed by more than 18 months of research on political disinformation campaigns. The GOST-ES prediction process has been used in research settings to develop a richer understanding of historical influence efforts, including respective target audiences, segments most likely to be impacted, and the influence efforts' impact on real-world outcomes such as public opinion and anti-Muslim violence.

GOST-ES is optimized to address three requirements:

1. Identify new or emerging foreign influence activities in near-real time across multiple platforms using content-based signatures from past activity, and active learning on real-time activity,
2. Protect privacy with minimal account-level information and no requirement for network or identifying data on specific users, and
3. Enable analysis of historical foreign influence operations by identifying:
   a. Trends in language usage and content promotion,
   b. Which kinds of activity set them apart from other users in each period, and
   c. Audience curation approaches.

Requirement One: Identifying Emerging Foreign Influence Activity

GOST-ES identifies new or emerging foreign influence activity in near-real time by learning from past activity across a range of platforms. It includes algorithms for feature generation, and models that rapidly identify new or emerging foreign influence via online social media. The system has been validated on multiple prediction challenges across four years of Chinese, Russian, and Venezuelan influence campaigns on Reddit and Twitter. The platform has a strong ability to detect activity from previously-unseen accounts that are part of known campaigns, is extremely robust to false negatives in training data (i.e. performance remains strong if only a small portion of past influence activity has been identified), and handles false positives well (i.e. if a few normal accounts are mislabeled as being part of an influence effort, it does not throw off the system).

GOST-ES captures data in a platform-agnostic manner on more than 1,000 features for every post, allowing it to provide rich data for other analytical workflows in support of communications and engagement activities. The system has already been applied to understand (i) the Internet Research Agency's production process, and (ii) the impact of Russian influence operations on real-world activity in the United States, including domestic terrorism. GOST-ES can ingest a wide range of publicly-available information (PAI) from the open and deep web, as well as multiple social media platforms, including Reddit, Twitter, 4chan, 8chan, and Line.

Requirement Two: Protect Privacy and Civil Liberties

GOST-ES does not rely on any account-specific information. It is designed to run on fully-anonymized data using only PAI. GOST-ES features are all based on aggregate reports, and output can be tailored to avoid identifying the specific classification of any particular post. GOST-ES explicitly models the behavior of normal politically-engaged users, helping to minimize false positives that might adversely impact the civil liberties of American citizens.

Requirement Three: Facilitate Analytics

GOST-ES produces a wide range of information on what distinguishes coordinated influence operation content from normal activity at any given point in time. In testing on past campaigns, we have found that the characteristic features of state-level political influence campaigns change dramatically over time as their tactics, techniques, and procedures change. Because GOST-ES algorithms explicitly track such changes, the modeling process creates extensive information that can be used for a range of analytics, including highlighting trending topics, calling out trending hashtags, identifying unusual n-grams, and surfacing key URLs that are promoted as part of an influence campaign. Together, this information allows near-real-time awareness of high-potential-impact foreign influence operations, narratives, and campaigns.