ROUTING AND ACTION

MEMORANDUM

ROUTING

TO:(1) Network Sciences Branch (Palazzolo, Edward)

Report is available for review

(2) Proposal Files Report No.: -DRP Proposal Number: 71737-NS-DRP.1

DESCRIPTION OF MATERIAL

CONTRACT OR GRANT NUMBER: W911NF-17-C-0095

INSTITUTION: Leidos, Inc.

PRINCIPAL INVESTIGATOR: Mark Williams

TYPE REPORT: Final Report

DATE RECEIVED: 1/31/22 9:15AM

PERIOD COVERED: 9/29/17 12:00AM through 12/31/21 12:00AM

TITLE: Final Report: Privacy-protected Environment for Gathering and Analyzing Social Information Spread (PEGASIS)

ACTION TAKEN BY DIVISION

(x) Report has been reviewed for technical sufficiency and IS [x] IS NOT [ ] satisfactory.

(x) Based on my technical review, I have identified no OPSEC or Technology Protection concerns that need to be addressed regarding this report.

(x) Perfomance of the research effort was accomplished in a satisfactory manner and all other technical requirements have been fulfilled.

(x) Based upon my knowledge of the research project, I agree with the patent information disclosed.

Approved by SSL\EDWARD.PALAZZOLO on 1/18/23 8:51PM

ARO FORM 36-E

REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
31-01-2022	Final Report	29-Sep-2017 - 31-Dec-2021

4. TITLE AND SUBTITLE

Final Report: Privacy-protected Environment for Gathering and Analyzing Social Information Spread (PEGASIS)

5a. CONTRACT NUMBER

5b. GRANT NUMBER

W911NF-17-C-0095

5c. PROGRAM ELEMENT NUMBER

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

Leidos, Inc.
1750 Presidents St

Reston, VA 20190 -5617

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)

ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

71737-NS-DRP.1

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Mark Williams
UU	UU	UU	UU		19b. TELEPHONE NUMBER 000-000-0000

RPPR Final Report
as of 18-Jan-2023

Agency Code: 21XD

Proposal Number: 71737NSDRP **Agreement Number: W911NF-17-C-0095**
INVESTIGATOR(S):

Name: Mark B. Williams
Email: mark.b.williams@leidos.com
Phone Number: 0000000000
Principal: Y

Organization: **Leidos, Inc.**
Address: 1750 Presidents St, Reston, VA 201905617
Country: USA
DUNS Number: 833063253 EIN: 95-3630868

Report Date: 31-Jan-2022 Date Received: 31-Jan-2022
Final Report for Period Beginning 29-Sep-2017 and Ending 31-Dec-2021
Title: Privacy-protected Environment for Gathering and Analyzing Social Information Spread (PEGASIS)
Begin Performance Period: 29-Sep-2017 **End Performance Period:** 31-Dec-2021
Report Term: 0-Other

Submitted By: Mark Williams Email: mark.b.williams@leidos.com
Phone: (000) 000-0000

STEM Degrees: 0 **STEM Participants:**

Major Goals: See PDF report for more details.

DARPA's Computational Simulation of Online Social Behavior program (SocialSim) sought to develop innovative technologies for high-fidelity computational simulation of online social behavior, focusing primarily on information spread and evolution within and across social media platforms. The Leidos team provided continuous and continually growing support to DARPA SocialSim from program inception in 2017 through the end of the program in 2022. Originally established as the Data Provisioning team, Leidos was responsible for data gathering and anonymization, data enrichment, gold standard ground truth delivery, and coordination with the Test and Evaluation team on challenge problem development. By mid-2019, Leidos assumed lead responsibility for challenge problem scenario development, to align the program more clearly with DoD missions and potential end-users. In mid-2020, our role on the program again expanded to include test and evaluation, the development of a visualization user interface, and the organization of outreach engagements with potential transition partners.

Accomplishments: See PDF report for more details.

Over the course of the program, the Leidos team was able to push the program forward through well-designed, operationally relevant challenge problems, through timely delivery of enriched, annotated, and privacy protected datasets, and eventually through comprehensive automated and human evaluation protocols which are detailed in the OEV final report. Our Privacy-protected Environment for Gathering and Analyzing Social Information Spread (PEGASIS) solution resulted in over 200 million online information environment (OIE) events across 6 platforms collected, anonymized, and enriched over the course of the program, with zero data spillage or privacy violations; and supported successful execution of six Challenge Problems (CPs) and publication of over 300 peer-reviewed papers.

Training Opportunities: Nothing to Report

RPPR Final Report
as of 18-Jan-2023

Results Dissemination: The Leidos team published the following papers:

Just Say No: Analyzing the Stance of Neural Dialogue Generation in Offensive Contexts
Ashutosh Baheti, Maarten Sap, Alan Ritter and Mark Riedl
Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 4846–4862

Neural semi-Markov CRF for Monolingual Word Alignment
Wuwei Lan *, Chao Jiang *, Wei Xu (* equal contribution)
Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL), pages 6815–6828

An Empirical Study of Pre-trained Transformers for Arabic Information Extraction
Wuwei Lan, Yang Chen, Wei Xu, Alan Ritter
Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 4727–4734

Corpus Development for Studying Online Disinformation Campaign: A Narrative + Stance Approach
Mack Blackburn, Ning Yu, John Berrie, Brian Gordon, David Longfellow, William Tirrell, Mark Williams
First International Workshop on Social Threats in Online Conversations, The 12th Language Resources and Evaluation Conference, pages 41-47

Detecting and Annotating Narratives in Social Media: A Vision Paper
Mack Blackburn, Ning Yu, Alex Memory, W. Graham Mueller
Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media

Code and Named Entity Recognition in StackOverflow
Jeniya Tabassum, Mounica Maddela, Wei Xu, Alan Ritter
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), pages 4913–4926

Measuring Forecasting Skill from Text
Shi Zong, Alan Ritter, Eduard Hovy
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), pages 5317–5331

Neural CRF Model for Sentence Alignment in Text Simplification
Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, Wei Xu
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), pages 7943–7960

Structured Minimally Supervised Learning for Neural Relation Extraction
Fan Bai and Alan Ritter
Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), pages 3057–3069

Analyzing the Perceived Severity of Cybersecurity Threats Reported on Social Media
Shi Zong, Alan Ritter, Graham Mueller, Evan Wright
Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), pages 1380–1390

Multi-task Pairwise Neural Ranking for Hashtag Segmentation
Mounica Maddela, Wei Xu, Daniel Preo?iuc-Pietro
Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), pages 2538–2549

Neural Network Models for Paraphrase Identification, Semantic Textual Similarity, Natural Language Inference, and Question Answering
Wuwei Lan, Wei Xu
Proceedings of the 27th International Conference on Computational Linguistics (COLING), pages 3890–3902, *Best Paper Award*

RPPR Final Report
as of 18-Jan-2023

Character-Based Neural Networks for Sentence Pair Modeling
Wuwei Lan, Wei Xu
Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL), pages 157–163

Honors and Awards: *Best Paper Award*:
"Neural Network Models for Paraphrase Identification, Semantic Textual Similarity, Natural Language Inference, and Question Answering"
Wuwei Lan, Wei Xu
Proceedings of the 27th International Conference on Computational Linguistics (COLING), pages 3890–3902

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: PD/PI
Participant: Mark Williams
Person Months Worked: 15.00
Project Contribution:
National Academy Member: N

Funding Support:

Partners

,

I certify that the information in the report is complete and accurate:
Signature: Mark Benjamin Williams
Signature Date: 1/31/22 9:15AM

W911NF-17-C-0095: DARPA SOCIALSIM Technical Area 2
Data Provisioning and Scenario Development

Final Report

Prepared by Leidos
Principal Investigator: Mark Williams
30 December 2021

This material is based upon work supported by, or in part by, the Army Research Laboratory and the Army Research Office under contract/grant number W911NF-17-C-0095.

ABSTRACT: Report developed under contract W911NF-17-C-0095. DARPA's Computational Simulation of Online Social Behavior program (SocialSim) sought to develop innovative technologies for high-fidelity computational simulation of online social behavior focusing primarily on information spread and evolution within and across social media platforms. We detail our work over the course of the program, collecting and provisioning enriched, privacy protected research corpora for training, testing, and evaluation. We also present methods to develop operationally relevant challenge problem scenarios and a scalable supervised approach to labeling narrative frames and stance positions over millions of social media events. The methods, enrichments, and approaches presented here can support efforts to better understand and counter online information/misinformation campaigns, support critical information delivery to local populations during HADR missions, and detect coordinated cyber threat activity early.

Table of Contents

1 Executive Summary

DARPA's Computational Simulation of Online Social Behavior program (SocialSim) sought to develop innovative technologies for high-fidelity computational simulation of online social behavior, focusing primarily on information spread and evolution within and across social media platforms. The Leidos team provided continuous and continually growing support to DARPA SocialSim from program inception in 2017 through the end of the program in 2022. Originally established as the Data Provisioning team, Leidos was responsible for data gathering and anonymization, data enrichment, gold standard ground truth delivery, and coordination with the Test and Evaluation team on challenge problem development. By mid-2019, Leidos assumed lead responsibility for challenge problem scenario development, to align the program more clearly with DoD missions and potential end-users. In mid-2020, our role on the program again expanded to include test and evaluation, the development of a visualization user interface, and the organization of outreach engagements with potential transition partners.[1]

Over the course of the program, the Leidos team was able to push the program forward through well-designed, operationally relevant challenge problems, through timely delivery of enriched, annotated, and privacy protected datasets, and eventually through comprehensive automated and human evaluation protocols which are detailed in the OEV final report. Our Privacy-protected Environment for Gathering and Analyzing Social Information Spread (PEGASIS) solution resulted in over 200 million online information environment (OIE) events across 6 platforms collected, anonymized, and enriched over the course of the program, with zero data spillage or privacy violations; and supported successful execution of six Challenge Problems (CPs) and publication of over 300 peer-reviewed papers.

This report covers the activities performed between September 2017 and December 2021.

[1] The Operational Evaluation and Visualization (OEV) work was performed under contract HR001120C0173 with DARPA.

2 Project Objectives

2.1 SocialSim Program Objectives

As stated in the Broad Agency Announcement, the Computational Simulation of Online Social Behavior (SocialSim) program goal is to develop innovative technologies for high-fidelity computational simulation of online social behavior. SocialSim will focus specifically on information spread and evolution. Specific SocialSim objectives are:

1. Develop technologies that can accurately simulate online information spread and evolution at scales representing populations of interest (i.e., thousands to millions).
2. Develop efficient and robust methods for providing data to support simulation development, testing, and measurement.
3. Develop rigorous methods and metrics for quantitatively assessing the accuracy and scalability of simulations of online information spread and evolution.

The Leidos team, comprised of engineers, data scientists, and OSINT practitioners, plus researchers Alan Ritter and Wei Xu at The Ohio State University and later Georgia Tech, and Kathleen Carley of Netanomics focused our TA2 Data Provisioning and TA3 Simulation Testing and Measurement efforts on objectives 2 and 3.

2.2 PEGASIS Project Objectives

The Leidos team extrapolated the above program goals into the following project objectives:

1. Data Gathering: Capture fast, complex, and often ephemeral information spread and evolution phenomena.
2. Data Gathering: Rapidly adapt to a changing online environment within and across multiple online information environment (OIE) platforms.
3. Privacy Protection: Protect personal privacy for online individuals without voiding data of analytic value.
4. Data Enrichment and Analysis: For meaningful analyses, assess the veracity of gathered data and improve quality as necessary.
5. Data Enrichment and Analysis: Identify and correlate events, topics, and specific messages across multiple information environments.
6. Data Enrichment and Analysis: Appropriately represent media content to enable scalable simulation of information spread and evolution.
7. Empirical Studies: Develop new instruments to study the causes of information spread and evolution in real or proxy environments.

3 Work Performed

3.1 Challenge Problem Design

The Leidos team coordinated with the original TA3 team from PNNL to conceptualize and design challenge problems that would require TA1 teams to develop novel approaches to simulating online information spread and social network activity. Together, we planned the baseline challenge and CPs 1-4. As the program evolved, Leidos assumed leadership in the design of challenges, from the selection of the overall context to the platforms, phenomena, and information components being studied.

As our role in CP design grew, we began coordinating stakeholder working groups to solicit input and buy-in from the operational community. Once an information domain had been chosen based on these discussions and conversations with DARPA, the TA1 and TA3 teams, Leidos conducted exploratory collection and analysis of several candidate scenarios before deciding on a final scenario. Finally, a collection strategy was defined specific to each CP.

3.1.1 Challenge Problem Roadmap

Challenge Problems 1 through 3 featured a focus on the spread of information associated with cyber vulnerabilities and/or cryptocurrency manipulation. Starting with Challenge Problem 3, we began focusing on contexts aligned with various information operations, from disinformation to censorship and soft influence operations. As the SocialSim Challenge Problem Roadmap [**Table 1**] shows, we varied the domains, scenarios, platforms, and volumes over the course of the program. This exposed TA1 teams to an array of information spread phenomena and use-case-specific characteristics and required them to adapt their modeling approaches to prevent overfitting.

Table 1: SocialSim Challenge Problem Roadmap

Domain	CP	Scenario	Platform	Social Media Events	Number of Metrics
Cyber	Baseline	GitHub Information Spread	GitHub	75 million	--
	CP1	CyberSec Vulnerabilities	GitHub / Twitter / Reddit	69 million	114
		Price Index	GitHub / Twitter / Reddit	24.5 million	
	CP2	State-Affiliated Exploits	GitHub / Twitter / Reddit	411K	204
		Pump 'n Dump	Twitter / Telegram	7.3 million	
	CP3	DOD-Relevant CVEs	GitHub / Twitter / Reddit	1.4 million	213
Information Operations		Syrian White Helmets	Twitter / YouTube	1.3 million	
	CP4	Venezuelan Presidential Crisis	Twitter / YouTube	40 million + 300K news articles	6
	CP5	Belt and Road in Pakistan	Twitter / YouTube / Reddit	3.7 million + 225K news articles	9 + Human Evaluation
	CP6	Belt and Road in East Africa	Twitter / YouTube / Reddit /Jamii	9 million + 515K news articles	14 + Human Evaluation
	Field Exp.	Hong Kong	Twitter	42 million + 86K news articles	6 + Human Evaluation

Working with the TA3 team from PNNL to refine methods and metrics for quantitatively assessing the accuracy and scalability of simulations, we also varied the number and specifics of the metrics used to evaluate models over the course of the program. In coordination with DARPA, Leidos observed by CP3 that there were too many metrics to clearly describe success, thus we reduced the number of metrics to focus more directly on assessing the accuracy of simulated volumes of shares and user activity. As in any requirements-driven system development effort, reducing the number of metrics encouraged TA1 teams to focus on and excel at a core set of "must-have" features in their model outputs, but it also allowed them to ignore other less essential output characteristics. For example, a model might correctly simulate the overall volume of social media posts on a given topic ("Total Shares" evaluated by Absolute Percentage Error), but they might inadvertently assign many topics to individual posts. In an effort to disincentivize these oversights, as the effective TA3 performer for CP5 forward, we incrementally added a discrete set of distinct and meaningful metrics and began performing an analyst-driven human evaluation to identify obviously unnatural outputs. See our OEV final report for additional discussion on the evolution of metrics.

3.1.2 Stakeholder Working Groups

Leidos coordinated and facilitated stakeholder working groups to solicit feedback and gain buy-in on the direction of the program and challenge problem design from the operational community. Participants varied from session to session, but often included representatives from Combatant Command J39s and J8s, members of the Intelligence Community, and analysts from warfighter support elements. Either as a break-out session during the periodic PI Meetings, or as virtual video conferences, these sessions frequently followed a similar agenda:

1. Orientation to SocialSim and Challenge Problem structure
2. Challenge Problem Roadmap
    a. Topics: identification of candidate scenarios
    b. OIE Platforms: discussion of which social media platforms were of top interest
    c. Phenomena: discussion of information spread phenomena of top interest
    d. Options for partnering
    e. Experimental Design: timing, scale, number of testing intervals, etc.
    f. Ground Truth elements: selection of `information IDs` to be tracked
    g. Measures and Metrics: identifying most meaningful elements to evaluate
3. Way forward for future engagement

3.1.3 Challenge Problem Exploration

Concurrent with the stakeholder working groups and coordination discussions with DARPA, and the TA1 and TA3 teams, Leidos and Netanomics analysts with backgrounds in information operations, intelligence analysis, cybersecurity, and cryptocurrency explored the feasibility of candidate scenarios. This activity included the specification of multi-lingual query terms, probes to determine the presence of suspected activity of interest on selected platforms, characterization of various forms of information spread and initial identification of candidate `information ID`s that could be further developed, tests of emerging data enrichment capabilities to assess utility for the given case, and ultimately a recommendation to proceed to

formal challenge problem selection or not. For more mature concepts, Leidos prepared overview slides detailing the description of the scenario, the operational relevance, data sources, phenomena under study, and general risks. An example slide is included as **Figure 1**.

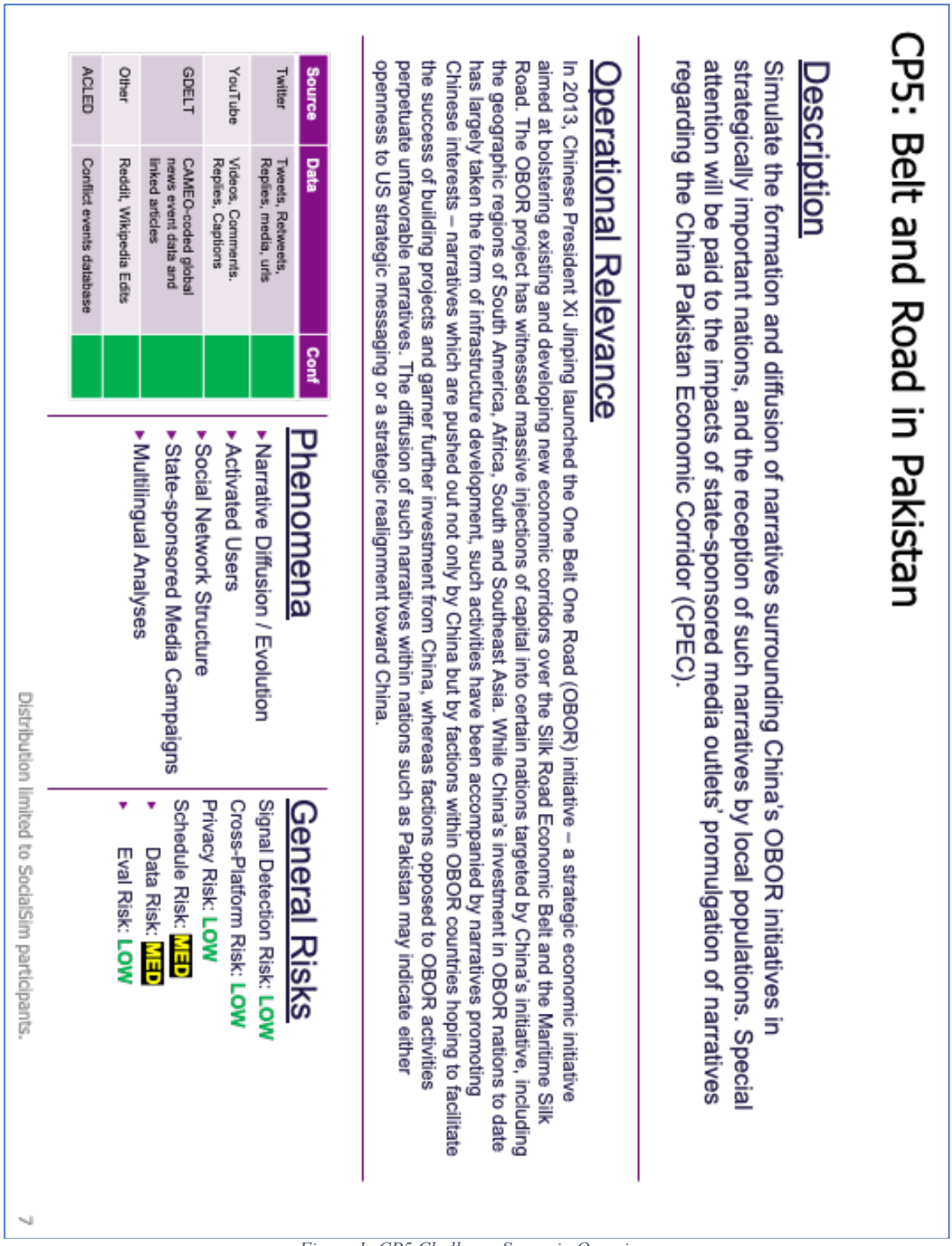

Figure 1: CP5 Challenge Scenario Overview

3.1.4 Collection Strategy Specification

Following selection of a given challenge problem scenario and endorsement by DARPA, Leidos formalized a collection and enrichment strategy specific to each use-case. As an example, for CP5 which focused on Twitter and YouTube used to discuss China's Belt and Road Initiative in Pakistan, we used the following query specifications and three-pass collection approach [**Figure 2**] to generate artifacts of cross-platform linkages and a fully networked sample of tweets:

Date Range:

- 30 March 2020 – 31 August 2020 (22 Weeks)

SME Curated Multi-lingual Keyword terms list:

- English: gwadar OR china-pak OR "china pakistan economic corridor" OR @cpec_official OR cpec OR cathaypak OR @pcipakchina OR "china-pakistan economic corridor" OR "gwadar port" OR cphgc OR pakchina_friendship OR ((pakistan OR baltistan OR baluchistan OR balochistan OR kashmir OR "Imran Khan" OR @ImranKhanPTI) AND ("Belt and Road Initiative" OR beltandroadpod OR OBORCHINA OR BeltandRoad OR BeltAndRoadInitiative OR #BRI OR #obor OR "one belt one road" OR "belt and road" OR beltandroad))
- Punjabi: ਚੀਨ AND (ਪਾਕਿਸਤਾਨ OR ਇਕ ਬੈਲਟ ਇਕ ਰੋਡ OR ਗਵਾਰਦਾਰ)
- Urdu: چین AND (ایک بیلٹ ون روڈ OR یو بیلٹ یو سڑک) OR پاکیستان

Three-pass Collection Strategy:

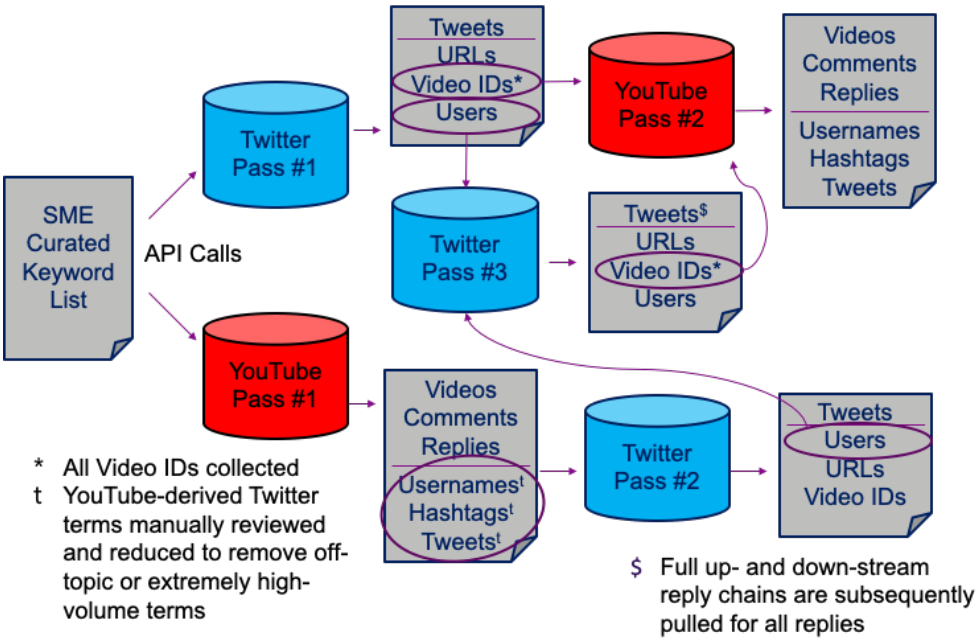

Figure 2: CP5 Data Collection Strategy

3.2 Collection

As defined through the Challenge Problem design and collection strategy specification, Leidos collected social media data to be developed into `ground truth` historical data for training and

testing of models, social media data to be used as non-modeled environmental information, and other open source and commercially available data to be used by the TA1 teams as "exogenous" environmental data that can inform simulations.

3.2.1 Data Working Groups

From program inception through Challenge Problem 3, Leidos lead regular Data Working Group engagements with the TA1 and TA3 performers to solicit feedback on enrichments, data structures, and candidate datasets. These sessions proved invaluable to both informing the other program performers about data and features that could be made available, and to help our team prioritize our efforts.

3.2.2 Online Information Environment (OIE) Data Collection

Leidos collected OIE data from the following platforms at various points in the program, both in direct support of challenge problems and to explore candidate scenarios and conduct experiments: Twitter, YouTube, Reddit, GitHub, Telegram, and Jamii Forums. **Table 2** details the social media content provided for each Challenge Problem.

Each platform presents its own challenges and opportunities for collection, and each brings a unique and ever-changing set of permitted (and prohibited) uses, terms of service, and APIs. We cultivated a positive, trust-based working relationship with Twitter, which enabled SocialSim researchers' access to larger and longer-duration samples than would be available via the public API. YouTube records are available via a public API, but as we show below, were almost impossible to collect a full history on. Historical Reddit records were conveniently made available to the entire community by an open-source blogger. GitHub records were retrievable via a series of APIs. IST Research made limited Telegram records available to the program for CP2. Lastly, we identified and collected data from the Jamii Forums, an East African message board-style forum for discussion of regional political issues, to test the generalizability of TA1 models to non-western social media platforms.

Table 2: Social Media data collected and provided to program participants.

CP	Scenario	Platform	Data Details
Baseline	GitHub Information Spread	GitHub	74,635,407 repos 20,657,898 users
CP1	CyberSec Vulnerabilities	GitHub	7,606,428 repos
		Twitter	30,704,015 Tweets
		Reddit	42,178,888 submissions and comments
	Price Index	GitHub	61,763 repos
		Twitter	18,432,161 Tweets
		Reddit	6,226,690 submissions and comments
CP2	State-Affiliated Exploits	GitHub	12,932 Repos 8,682 unique users 1,165 exploited CVEs
		Twitter	93,928 Tweets
		Reddit	3,064 relevant comments + complete thread

			1,718 relevant submissions + complete thread 1,400 relevant submissions in CVEWatch subreddit
	Pump ‘n Dump	Twitter	5,239,402 Tweets
		Telegram	39 channels/groups 60 Users 403,739 messages
CP3	DOD-Relevant CVEs	GitHub	17,135 repos 216,267 events 26,424 users
		Twitter	523,130 Tweets
		Reddit	2,480 post authors 116,481 comment authors 519,235 submissions and comments
	Syrian White Helmets	Twitter	1,102,007 Tweets
		YouTube	1,753 Videos 631 Channels 78,762 Comments
CP4	Venezuelan Presidential Crisis	Twitter	27,888,731 Tweets
		YouTube	10,974 Videos 4,427 Channels 3,192,169 Comments
CP5	Belt and Road in Pakistan	Twitter	2,098,527 Tweets
		YouTube	4,943 Videos 2,625 Channels 432,805 Comments
		Reddit	33,371 Posts 746,319 Comments
CP6	Belt and Road in East Africa	Twitter	6,117,493 Tweets
		YouTube	15,876 Videos 3,093 Channels 430,696 Comments
		Reddit	19,891 Posts 901,659 Comments
		Jamii	5,867 Posts 413,223 Comments
Field Exp.	Hong Kong	Twitter	41,445,960 Tweets

3.2.3 Exogenous Data Collection

Whereas social media data was collected and processed to develop a historical ground truth dataset to be used by the TA1 teams to train and test their simulators, we also collected and provided large volumes of exogenous data that could optionally be leveraged by the performers to inform their modeling approaches and improve the accuracy of their simulations. As conversations on social media often parallel and intersect with current events as covered in the traditional news media, some of the most powerful sources of exogenous data were aggregations of news-worthy events and raw, full-text digital news articles. We also provided

economic data such as currency exchange rates (crypto conversion rates), commodity price histories (oil prices), curated event histories (violence, internet outages, electoral timelines), and other online user-generated content (Wikipedia edits, dark web activity). Whenever possible, we would provide translations and native-language versions of texts. **Table 3** details the exogenous data provided for each Challenge Problem.

Table 3: Exogenous data collected and provided to program participants.

CP	Scenario	Data Type	Data Details
CP1	CyberSec Vulnerabilities	Cyber Events	Hackmageddon list of major cyber incidents
		HackerNews	Compendium of cyber vulnerability news sources
		NVD	National Vulnerability Database
	Price Index	Crypto Price	Daily log of USD-equivalent prices for each cryptocoin
CP2	State-Affiliated Exploits	DarkOwl	Dark Web mention stats of SocialSim topics/keywords 71,980 articles
		HackerNews	Compendium of cyber vulnerability news sources
	Pump 'n Dump	Crypto Price	Daily log of USD-equivalent prices for each cryptocoin
		DarkOwl	Dark Web mention stats of SocialSim topics/keywords 53,571 articles
CP3	DOD-Relevant CVEs	MITRE CVE list	Comprehensive list of CVEs and their details
		NVD Vendor list	National Vulnerability Database down-selected to known DoD vendors
	Syrian White Helmets	GDELT	2,848,170 machine coded geopolitical events
CP4	Venezuelan Presidential Crisis	GDELT	639,915 machine coded geopolitical events
		Digital News	146,619 news articles linked from GDELT events and/or the Social Media corpus
		Oil Prices	West Texas Intermediate and Brent Crude daily spot prices through 2019
		Internet Outages	Calendar of 54 internet outages reported in Venezuela from NetBlocks.org
CP5	Belt and Road in Pakistan	GDELT	1 million machine coded geopolitical events
		ACLED	3,993 human-reviewed armed conflict events in Pakistan
		Wikipedia Revisions	200 articles and 12,902 revisions to pages referenced in the Social Media corpus
		Digital News	172,396 news articles linked from GDELT events and/or the Social Media corpus
CP6	Belt and Road in East Africa	GDELT	1.7 million machine coded geopolitical events
		ACLED	3,600 human-reviewed armed conflict events in East Africa
		Digital News	480,000 news articles linked from GDELT events and/or the Social Media corpus
		LORELEI	30,000 extracted events from the DARPA LORELEI program
Field Exp.	Hong Kong	GDELT	6,493 machine coded geopolitical events
		Digital News	85,365 news articles linked from GDELT events and/or the Social Media corpus

3.3 Privacy Protection

The protection of online users' privacy was of principal import, both to satisfy IRB/HSR requirements and to gain the trust of data providers in order to share large volumes of user-generated content with program participants.

Leidos Institutional Review Board (IRB) reviewed our SocialSim proposal and reviewed the project in execution. Based on the absence of direct human subjects research and the inclusion of extensive privacy protection protocols, Leidos declared the project as exempt from further IRB oversight. DARPA HRPO also completed its Government review of the program and our work and privacy safeguards, determining that we would be exempt from further HSR reviews. Our research design and data handling protocols ensured information recorded by the investigator(s) is performed in such a manner that subjects cannot be identified, directly or indirectly through identifiers linked to the subjects. Specifically included in our risk reduction protocols were the following:

- For each new data set or data source, manually verify the ability of automated technical means to eliminate PII.
- Partition any data found retaining PII after release. Immediately restrict access and require program performers to destroy any local copies of the data pending Leidos and DARPA investigation of root cause and removal of PII.
- Raw collected data will be removed from the SocialSim DMZ and destroyed at the end of every phase in compliance with the SocialSim Privacy Plan.
- Require SocialSim Privacy training (DARPA) for researchers.

Online information providers ultimately own and control the data generated on their platforms. Many of these companies have built successful brands and loyal userbases based in part on a publicly reinforced reputation of protecting their customers' privacy. Twitter, for instance, states in their Developer Agreement and Policy: "Privacy and Control are Essential - Protecting and defending the privacy of people on Twitter is built into the core DNA of our company. As such, we prohibit the use of Twitter data in any way that would be inconsistent with people's reasonable expectations of privacy."[2] Further, Twitter's terms limit the sharing of Tweet IDs (the unique identifier of a tweet) to 1.5 million within a 30-day period and not more than 50,000 Tweet Objects (the details and content) without written permission from Twitter. We obtained that written permission for explicitly defined challenge problem scenarios based on the strength of our privacy protection and anonymization protocols. Given that we often needed to share tens of millions of Tweet Objects with TA1 and TA3 teams, the program would not have been able to use Twitter data otherwise.

[2] Twitter Developer Terms: Developer Agreement and Policy, effective March 10, 2020. https://developer.twitter.com/en/developer-terms/agreement-and-policy

3.3.1 Privacy Plan

Leidos follows the privacy protection plan developed by DARPA I2O for the SocialSim program. Among other requirements, the SocialSim Privacy Plan mandates that all PII data "processed by the program will be:

- Encrypted
- Temporarily stored on a secure server; firewall protected
- As feasible and effective, processed with automated tools that can eliminate sensitive information
- Access-limited to designated privacy-trained program personnel for program purposes only
- Not shared outside of DARPA except in accordance with the program's sharing procedure
- Destroyed at completion of the program."[3]

3.3.2 Data Use Agreements

Leidos established Limited Access and Non-Disclosure Agreements with all SocialSim performer teams, the TA3 evaluator, and affiliated SBIR performers at the direction of the SocialSim Program Manager. This agreement requires teams receiving the huge amounts of potentially privacy-impacting data to protect that data, designate approved and appropriately trained users, limit their use of the data to SocialSim research, and delete the data upon program completion.

Leidos established a data use agreement with Twitter that allows Leidos to collect data through Twitter's PowerTrack API. Advantages of collecting data through the PowerTrack API include larger volumes, collecting historical data, and more metadata than through the public API. As part of its data use agreement, Leidos and Twitter collaboratively defined and scoped the scenarios before initiating data set development. This process is multi-step. Leidos prepared a brief plain-language summary for each candidate scenario of interest. Leidos performed initial collection and analysis of the data to extract topics and hashtags associated with the scenarios. Twitter reviewed, provided feedback, and ultimately approved/rejected the scenario request from Leidos. When approved, Leidos initiated the full collection and developed the data set, performed the privacy protection routines detailed below, and distributed anonymized copies to the approved performers. Periodically, Twitter conducts an independent review of all parties who have access to their data through the master Leidos data use agreement. As a result of this review, Twitter occasionally revokes access to its data as a result of these reviews, but does not share the details of their independent assessments.

Leidos has been able to negotiate extended access to data collected under the Social Sim program for participants under the existing data use agreements. Notably, with DARPA endorsement, we are able to extend current access to collected and anonymized Twitter data through August 2022, and have maintained the ability to perform additional collections

[3] SocialSim Program Privacy Plan, DARPA I2O/MSO/GC, Distribution A

supporting other DARPA programs in the Confidence in the Information Domain strategic thrust through August 2022, when our license expires. With renewed funding, this capability could be further extended beyond August 2022. Additionally, Leidos has prepared redacted datasets that maintain referential integrity between Social Sim enrichments and TweetIDs, that are able to persist in perpetuity to support future academic research.

3.3.3 Privacy Implementation

Leidos leveraged its experience developing anonymized data sets for the DoD and Intelligence Community to support research, development, and evaluation activities for DARPA SocialSim. Leidos has developed deep capabilities and experience enforcing and following an extensively reviewed and approved DoD 5240.1-R and DoD Manual 5240.01 compliant process to identify and restrict access to sensitive U.S. persons data in large collections through our Intelligence Community programs.

3.3.3.1 Data collection and storage

Leidos downloaded collected real-world data directly into a demilitarized zone (DMZ). The DMZ has a double firewall designed to block external access and unauthorized internal access: the front-end firewall is configured to allow traffic destined to the DMZ only, and the back-end firewall only allows traffic from the DMZ to the internal network. Internally enforced requirements include:

- Perform virus scanning on the raw data and any data transformation,
- Encrypt data in all electronic transmission,
- Signed nondisclosure agreements for all users,
- Log and document all access to the data to show person, time, and nature of data access, and
- Destroy data after required reviews and mandatory retention periods.

The Leidos VPN provides key-based, role-oriented authentication and authorization services. Authenticated users (Leidos) access the system to review data at levels appropriate to assigned roles. Researchers have access only to deidentified, anonymized data, not the raw data.

Collected data reside in secure Elasticsearch indexes, partitioned in a DMZ that is exposed only to Leidos. PII removal tools transform raw data in these secure indices into consistent, anonymized indexes. Once data are cleaned they are exposed from within the DMZ for access by analytics services on the Leidos VPN.

3.3.3.2 PII Identification

Leidos applies semi-automatic methods to identify and remove direct and indirect identifiers. A direct identifier is information that can unequivocally and by itself identify the person in question. Examples for direct identifiers include, for example, names (e.g., profile name, screen name), address, home page, or contact information (e.g., phone, email, WeChat id). An indirect identifier is information that combines with other information to identify an individual, even though that individual piece of information is harmless by itself – similar to the mosaic concern in national security. Examples for indirect identifiers include, for example, user activities (e.g.,

time a message was posted), general location (restaurant, city), and socio-demographic information, such as income or birth date.

Direct and indirect identifiers may appear as metadata or within messages themselves and may vary by online platform. For example, Facebook provides a city attribute for its user; Twitter does not. Leidos lists sensitive information fields for each target platform and keeps these lists up-to-date. For identifiers mentioned in text, Leidos uses a list of pattern-based extractors, and removes, for example, sequences of eight or more digits that might correspond to credit card or social security numbers. This list combines patterns Leidos teams have developed for other research programs, including 1) extractors we developed for the DARPA MEMEX program to identify names, phone numbers, messenger IDs, and emails; and 2) new capabilities we developed under SocialSim to prepare Twitter data.

3.3.3.3 PII Removal

Direct Identifier Removal: After identifying the direct identifiers, Leidos conducts a selective class-based masking approach to decide whether to remove direct identifiers based on the entity class (e.g., user, celebrity, organization, news agency). For example, the name of a news agency would not be masked, but the name of an individual would be de-identified by replacing the direct identifier with a unique hash alias. Leidos uses MD5 hash generator for generating the unique hash aliases.

Indirect Identifier Removal: Because many identifiers have useful properties for simulation development, handling indirect identifiers is more complicated than direct identifiers. For example, to conduct multi-resolution simulation, researchers may group actors by geolocation, gender, and other properties.

3.4 Enrichments

In addition to collecting and anonymizing social media data, Leidos and our teammates delivered a substantial number of enrichments along with each dataset to further enable the TA1 research teams to understand and therefore more accurately simulate information spread. These enrichments ranged from straightforward, but computationally intensive processes better performed once by the TA2 team than by each of the TA1 teams (i.e. machine translation, languageID), to external API queries requiring the non-anonymized data for input (i.e. botometer, bothunter), to bespoke models to tag some feature of interest (i.e. supervised narrative labeling models). Throughout the program, we periodically surveyed the TA1 performers to understand which enrichments were most valuable, which additional enrichments they would like to see, and which were not adding additional value but might have been delaying data delivery.

In most cases, we were able to provide desired enrichments or exogenous data. One notable case where we were not able to provide desirable enrichments was regarding lists (or counts) of Twitter “followers.” Unfortunately, the list (and count) of a given Twitter user’s followers is only available as of the moment that collection is executed, not as of the moment that a tweet is sent. Thus, the list or count of followers we could generate would not be contemporaneous

to the messages and replies we were collecting historically. The only workaround would be to collect against a scenario in real-time, populating a database of followers linked to each tweet or reply.

Figure 3 shows our end-to-end pipeline concept for data ingest, enrichment, and ground truth generation. Notice that at times, the outputs from enrichment processes would be used to inform secondary collection steps, such as when we would use our "on topic" content labeler to down-select users for follow-on collection (see **Figure 2**).

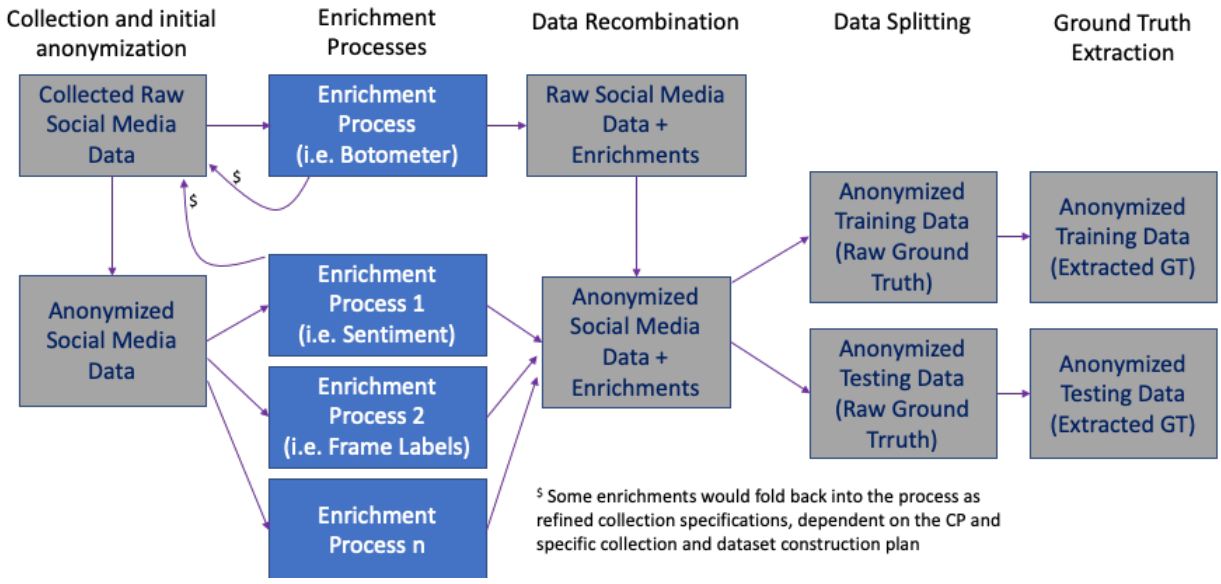

Figure 3: Data Collection, Enrichment, and Ground Truth Generation Pipeline

3.4.1 Information ID Labeling

Beginning with the Syrian White Helmets scenario for Challenge Problem 3, the Leidos team implemented a supervised learning approach to create `information ID` labels for social media events that would be simulated by the TA1 teams. Whereas in earlier challenges, TA1 teams were asked to simulate the spread of discrete `information ID`s, such as hashtags, URLs, CVE numbers, and other well-formed text strings, the new Information Operations use-cases required simulating the spread of more nebulous concepts such as stances on issues and narrative frames. In order to develop high-quality labels that could be scale to the millions of social media events in our corpora, we developed a multi-step human annotation and machine learning approach, which was repeated and refined from CP3 through the field experiment following CP6.

In each case, groups of analysts with subject-matter familiarity would convene to identify likely narrative frames and stance values that would be found in our collections. We would perform unsupervised topic modeling over the corpus to help guide this conversation. Following the development of the initial list of labels and definitions, the analyst-annotators began labeling small samples of social media events. Working through multiple iterations, with periodic inter-

rater reliability scoring, and follow-up meetings to discuss discrepancies and refine the annotation guidelines, the team would reach consensus on a set of annotation guidelines and begin marking up thousands of events to create a training set. Inter-rater reliability (IRR) would again be performed on events in common, and the labels with lower IRR scores or extremely low numbers of exemplars would be removed from the training set. Multiple language models would be trained and evaluated, and the best performing model (as evaluated against the humans) would be selected to generate labels for the entire challenge problem corpus. Annotation Guidelines for CP4-CP6 are included in the Annex.

Table 4 details many of the enrichments provided throughout the program.

Table 4: Enrichments Provided throughout SocialSim

CP	Platform	Enrichments
CP1	Twitter	User level –Organization VS. individual –Bots –Gender Tweet level –Sentiment –Targeted sentiment –Hashtag segment E.g., #ilovebitcoin -> I love bitcoin –Tweet country location –Politeness and psycholinguistic features(SIFT) –Reaction between quote tweet and reply –Image dhash –Uncertainty –NER Network level –Twitter followers –Twitter specific network metrics E.g., basic groups and stance, influentialness, opinion leader, super spreader, super friends, group identification Retweet network estimation
CP1	Reddit	Sentiment label (subjective, positive, negative) Targeted sentiment Perceptual image hashing: dhash Uncertainty – example: “Perhaps my account has been attacked”. Label: uncertain, Score: 0.52 • Politeness and psycholinguistic features (SIFT) • Reaction
CP2	Twitter	domain-specific tokenization sentiment hashtag segmentation named entity recognition languageID, Sentiment

CP2	GitHub	domain-specific tokenization pre-trained word embeddings languageID
CP3	Twitter	extracted_ner segmented_hashtags predicted_sentiment sentiment_scores narrative tagging botometer Retweet Network Reconstruction Topic Vector Representation Machine Translation of Arabic Tweets
CP3	GitHub	domain-specific tokenization pre-trained word embeddings languageID
CP4	Twitter	Botometer score for posting user Bothunter score for the posting user Humanizr classification for posting user Sentiment of tweet text Retweet Reconstruction Translation of tweet text with Google Translate Resolved URLs Narrative
CP4	YouTube	Detection of Language with PolyGlot Translation of Text with Google Translate Sentiment Polarity and Subjectivity with PloyGlot (English) or TextBlob (non-English) Urls from text extracted and resolved.
CP5	Twitter	Manual Frame/Stance/Antagonist/Protagonist: From annotation effort Supervised Frame/Stance Topic Vector (en) Stanford OpenIE subject-relation-object triples Google Translation of non-English Botometer BotHunter Humanizr Sentiment Named Entity Extraction (NER) Hashtag Segmentation Retweet Reconstruction
CP5	YouTube	Manual Frame/Stance/Antagonist/Protagonist: From annotation effort Supervised Frame/Stance Topic Vector (en) Stanford OpenIE subject-relation-object triples Google Translation of non-English
CP6	Twitter / YouTube / Jamii / Reddit	Supervised Frame Supervised Stance Supervised On Topic Manual Annotation Labels Google Translation (if not detected English) With Language Detection OpenIE Relation Extraction on translated text if not English Topic Modelling on translated text if not English

		Search Keywords Matched PETRARCH2 CAMEO Event and Issue Codes on translated text if not English Twitter Only: - Botometer User Classification - BotHunter User Classification - Humanizr Classification - Hashtag Segmentation - Sentiment - NER

3.5 Data Delivery

We provide secure data delivery for the TA3 and TA1 program participants. Our solution satisfies compliance requirements from the data use agreement and privacy plan, multiple data delivery use cases with different user cohorts, complex batches of data with several large files, evolving participants and use cases and time sensitive data releases to support challenge problem schedules. We designed a mutli-faceted cloud-based solution using AWS. In this section we detail the policies, use cases and infrastructure underpinning the solution.

3.5.1 Provision of AWS environment to host data and webservice.

Our solution architecture includes a way to deliver data with 3 classes of publicity:

1. All members of the data access group, TA1, TA2 and TA3. These releases are documented on the SocialSim Wiki and linked from there.
2. Private coordination with TA2 remote partners for enrichment and exploration tasks.
3. Private coordination with TA3 to support challenge problem development and evaluation.

Only members who satisfy the data use agreement have access to the environment. By default that applies them to group 1. If they are under the TA2 task, they are also added to group 2. Any TA3 participants are added to group 3.

Group 1 releases are public to the program and are published additionally by a Web Service. This allows them to be explored and downloaded via the browser. This simplified linking the releases from the wiki and access complexity by all range of participants. To support technical network to network transfers, we also granted SSH tunnel access to the deployment directories and published rsync commands and procedures to allow technical teams to keep their local environments up to date with the official release contents.

Group 2 and 3 have different deployment directories and only SSH tunnel access. Conceptually they are kept apart from both Group 1 and each other, but all members are in Group 1.

This architecture is realized with a user directory with group management (LDAP), which is updated by the data use agreement procedures. A VPN then governs 2-factor access via proxy to the environment. An EC2 with a large data volume supports the HTTP server and group-level access-controlled directories. Group level access is managed by access to the LDAP server.

3.5.2 VPN with 2FA access control tied to schedule A of data use agreement.

The data delivery environment user list is updated through governance via the data use agreement procedures (Section LINKME). When a user is added to a performer's Schedule A of the data use agreement, they are added to the user list with proper group access based on the organization's data use agreement. We retained and reviewed the master user list as project personnel changed.

We configured the SocialSim VPN for username, password and PKI to satisfy 2-factor authentication requirements. All users are given accounts with one time passwords with certificates. Both are needed to authenticate with the VPN sever. The same username and password are used to access the data repositories wither via HTTP or rsync once connected to the network.

3.5.3 Delivery of event and exogenous data files

We release data through a copying of data from PegasisNet to one of the 3 group destinations. PegasisNet contains staging areas for release where final checks are done to confirm compliance and final QC checks. Once everything passes, we finalize the delivery by producing a checksum and use rsync to replicate the staging directory to the proper release location on the delivery environment. If it is a Group 1 release, we document the release with detailed descriptions and links on the SocialSim Wiki. Wiki page updates act as the publication of the release. For Group 2 and 3 releases, we include a readme describing the release and notify the recipient via email.

Release contents and formats evolved as part of an ongoing process with the consumers. The main format was json-lines described by the filename using directories to sort via type and source. The further structure is filled out by the documentation on the SocialSim Wiki. We also included ElasticSearch indexed data for Challenge Problems 2, 3 and 4. We added an ElasticSearch cluster within the environment, an S3 bucket for capturing snapshots and publishing snapshots for each release. This streamlined ingest to a new ElasticSearch environment by TA1 performers. Over time, the need for this level of integration was deprioritized and we dropped this support for challenges 5 and 6. We were able to downsize the environment to save on cost.

As version tracking became increasingly important, we designed versioning schemes to improve provenance of what was changing with a new version, what is the current version, and how to identify what version you are using. We leveraged different combinations of naming conventions, detailed SocialSim Wiki documentation highlighting the provenance of all versions, and always having a current version checksum on all releases. We used the SocialSim Wiki release pages to field questions on the data releases, allowing for all members to see the discussion and further support the provenance of releases.

4 Results Obtained

We were able to make demonstrable progress in the development of a number of textual enrichments, and online information environment classification tasks that became integral to the execution of the program.

4.1 Research Insights and Novel Ideas

The principal work to deliver a methodology and customized algorithms to label millions of social media posts with narrative, narrative frame, and stance labels transformed the ability to track the spread of abstract information concepts rather than just string-matched text snippets such as hashtags or URLs. That work was detailed above and in Blackburn et al.'s "Corpus Development for Studying Online Disinformation Campaign: A Narrative + Stance Approach" and "Detecting and Annotating Narratives in Social Media: A Vision Paper." The resultant language models have proven to be able to be applied against independently collected corpora, including both long-form traditional news media full text and short-form news media headlines. The methodology has also shown to be reliably repeatable for alternative scenario-based collections, regardless of language or source.

Our teammates at Ohio State University and later Georgia Tech made several science contributions and technology innovations. To support data enrichment for Twitter, we developed a new tool (HashtagMaster) for analyzing hashtags that uses a novel pairwise neural ranking model and can convert hashtags into a meaningful word sequence (e.g., #EarthquakeinKirinyaga -> Earthquake in Kirinyaga) with a very high accuracy (>90%). We created specialized pre-trained language models and entity recognition models for the StackOverflow domain, in particular to support SocialSim program's initial goal in Year 1. This work has been increasingly used and cited by the software engineer research community. We also developed semantic models for capturing meaning-equivalences (i.e., paraphrases) that convey the same meaning with variants in word choices, which is crucial for tracking information flow in social media. Our work won a paper award at COLING 2018. Work on toxic language models trained on Reddit data found that neural chatbots are twice as likely to agree with offensive comments, which appears to be due to a filter-bubble effect in the internet conversations these models are trained on. This work was covered by TNW.[4] We analyzed the language used by people who make predictions about the future in online forums and found skilled forecasters tend to hedge their statements with more uncertainty and also make more references to past events. This supports existing theories about human forecasting skill in the psychology literature. Lastly, we developed a system that extracts early reports of software vulnerabilities mentioned on Twitter, and analyzes users' opinions about their severity [8]. By comparing users' opinions against expert judgments in the National Vulnerability Database, we showed online opinions provide a potentially useful early indicator of severe vulnerabilities. This work was covered by WIRED.[5]

[4] https://thenextweb.com/news/gpt-3-and-humans-twice-as-likely-agree-with-offensive-reddit-comments-chatbots/

[5] https://www.wired.com/story/machine-learning-tweets-critical-security-flaws/

4.2 Published Papers

While not a primary objective of the TA2 work, the Leidos team was able to publish the following papers detailing results of our research and related studies:

Just Say No: Analyzing the Stance of Neural Dialogue Generation in Offensive Contexts
Ashutosh Baheti, Maarten Sap, Alan Ritter and Mark Riedl
Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 4846–4862
Abstract: Dialogue models trained on human conversations inadvertently learn to generate toxic responses. In addition to producing explicitly offensive utterances, these models can also implicitly insult a group or individual by aligning themselves with an offensive statement. To better understand the dynamics of contextually offensive language, we investigate the stance of dialogue model responses in offensive Reddit conversations.

Neural semi-Markov CRF for Monolingual Word Alignment
Wuwei Lan *, Chao Jiang *, Wei Xu (* equal contribution)
Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL), pages 6815–6828
Abstract: Monolingual word alignment is important for studying fine-grained editing operations (i.e., deletion, addition, and substitution) in text-to-text generation tasks, such as paraphrase generation, text simplification, neutralizing biased language, etc. In this paper, we present a novel neural semi-Markov CRF alignment model, which unifies word and phrase alignments through variable-length spans. We also create a new benchmark with human annotations that cover four different text genres to evaluate monolingual word alignment models in more realistic settings. Experimental results show that our proposed model outperforms all previous approaches for monolingual word alignment as well as a competitive QA-based baseline, which was previously only applied to bilingual data. Our model demonstrates good generalizability to three out-of-domain datasets and shows great utility in two downstream applications: automatic text simplification and sentence pair classification tasks.

An Empirical Study of Pre-trained Transformers for Arabic Information Extraction
Wuwei Lan, Yang Chen, Wei Xu, Alan Ritter
Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 4727–4734
Abstract: Multilingual pre-trained Transformers, such as mBERT (Devlin et al., 2019) and XLM-RoBERTa (Conneau et al., 2020a), have been shown to enable the effective cross-lingual zero-shot transfer. However, their performance on Arabic information extraction (IE) tasks is not very well studied. In this paper, we pre-train a customized bilingual BERT, dubbed GigaBERT, that is designed specifically for Arabic NLP and English-to-Arabic zero-shot transfer learning. We study GigaBERT's effectiveness on zero-short transfer across four IE tasks: named entity recognition, part-of-speech tagging, argument role labeling, and relation extraction. Our best

model significantly outperforms mBERT, XLM-RoBERTa, and AraBERT (Antoun et al., 2020) in both the supervised and zero-shot transfer settings.

Corpus Development for Studying Online Disinformation Campaign: A Narrative + Stance Approach
Mack Blackburn, Ning Yu, John Berrie, Brian Gordon, David Longfellow, William Tirrell, Mark Williams
First International Workshop on Social Threats in Online Conversations, The 12th Language Resources and Evaluation Conference, pages 41-47
Abstract: Disinformation on social media is impacting our personal life and society. The outbreak of the new coronavirus is the most recent example for which a wealth of disinformation provoked fear, hate, and even social panic. While there are emerging interests in studying how disinformation campaigns form, spread, and influence target audiences, developing disinformation campaign corpora is challenging given the high volume, fast evolution, and wide variation of messages associated with each campaign. Disinformation cannot always be captured by simple factchecking, which makes it even more challenging to validate and create ground truth. This paper presents our approach to develop a corpus for studying disinformation campaigns targeting the White Helmets of Syria. We bypass directly classifying a piece of information as disinformation or not. Instead, we label the narrative and stance of tweets and YouTube comments about White Helmets.

Detecting and Annotating Narratives in Social Media: A Vision Paper
Mack Blackburn, Ning Yu, Alex Memory, W. Graham Mueller
Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media
Abstract: Understanding how narratives spread and influence target audiences in dynamic information environments like social media has a broad range of applications, e.g., to anticipate public reaction to policy changes or disinformation campaigns. In this paper we describe our ongoing efforts to define, collect, detect, and annotate narratives online in order to analyze and model the spread of these narratives. We also share lessons learned and our vision for how to improve these sub-tasks.

Code and Named Entity Recognition in StackOverflow
Jeniya Tabassum, Mounica Maddela, Wei Xu, Alan Ritter
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), pages 4913–4926
Abstract: There is an increasing interest in studying natural language and computer code together, as large corpora of programming texts become readily available on the Internet. For example, StackOverflow currently has over 15 million programming related questions written by 8.5 million users. Meanwhile, there is still a lack of fundamental NLP techniques for identifying code tokens or software-related named entities that appear within natural language sentences. In this paper, we introduce a new named entity recognition (NER) corpus for the

computer programming domain, consisting of 15,372 sentences annotated with 20 fine-grained entity types.

Measuring Forecasting Skill from Text
Shi Zong, Alan Ritter, Eduard Hovy
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), pages 5317–5331
Abstract: People vary in their ability to make accurate predictions about the future. Prior studies have shown that some individuals can predict the outcome of future events with consistently better accuracy. This leads to a natural question: what makes some forecasters better than others? In this paper we explore connections between the language people use to describe their predictions and their forecasting skill. Datasets from two different forecasting domains are explored: (1) geopolitical forecasts from Good Judgment Open, an online prediction forum and (2) a corpus of company earnings forecasts made by financial analysts. We present a number of linguistic metrics which are computed over text associated with people's predictions about the future including: uncertainty, readability, and emotion. By studying linguistic factors associated with predictions, we are able to shed some light on the approach taken by skilled forecasters. Furthermore, we demonstrate that it is possible to accurately predict forecasting skill using a model that is based solely on language. This could potentially be useful for identifying accurate predictions or potentially skilled forecasters earlier.

Neural CRF Model for Sentence Alignment in Text Simplification
Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, Wei Xu
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), pages 7943–7960
Abstract: The success of a text simplification system heavily depends on the quality and quantity of complex-simple sentence pairs in the training corpus, which are extracted by aligning sentences between parallel articles. To evaluate and improve sentence alignment quality, we create two manually annotated sentence-aligned datasets from two commonly used text simplification corpora, Newsela and Wikipedia. We propose a novel neural CRF alignment model which not only leverages the sequential nature of sentences in parallel documents but also utilizes a neural sentence pair model to capture semantic similarity. Experiments demonstrate that our proposed approach outperforms all the previous work on monolingual sentence alignment task by more than 5 points in F1. We apply our CRF aligner to construct two new text simplification datasets, Newsela-Auto and Wiki-Auto, which are much larger and of better quality compared to the existing datasets. A Transformer-based seq2seq model trained on our datasets establishes a new state-of-the-art for text simplification in both automatic and human evaluation.

Structured Minimally Supervised Learning for Neural Relation Extraction
Fan Bai and Alan Ritter

Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), pages 3057–3069
Abstract: We present an approach to minimally supervised relation extraction that combines the benefits of learned representations and structured learning, and accurately predicts sentence-level relation mentions given only proposition-level supervision from a KB. By explicitly reasoning about missing data during learning, our approach enables large-scale training of 1D convolutional neural networks while mitigating the issue of label noise inherent in distant supervision. Our approach achieves state-of-the-art results on minimally supervised sentential relation extraction, outperforming a number of baselines, including a competitive approach that uses the attention layer of a purely neural model.

Analyzing the Perceived Severity of Cybersecurity Threats Reported on Social Media
Shi Zong, Alan Ritter, Graham Mueller, Evan Wright
Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), pages 1380–1390
Abstract: Breaking cybersecurity events are shared across a range of websites, including security blogs (FireEye, Kaspersky, etc.), in addition to social media platforms such as Facebook and Twitter. In this paper, we investigate methods to analyze the severity of cybersecurity threats based on the language that is used to describe them online. A corpus of 6,000 tweets describing software vulnerabilities is annotated with authors' opinions toward their severity. We show that our corpus supports the development of automatic classifiers with high precision for this task. Furthermore, we demonstrate the value of analyzing users' opinions about the severity of threats reported online as an early indicator of important software vulnerabilities. We present a simple, yet effective method for linking software vulnerabilities reported in tweets to Common Vulnerabilities and Exposures (CVEs) in the National Vulnerability Database (NVD). Using our predicted severity scores, we show that it is possible to achieve a Precision@50 of 0.86 when forecasting high severity vulnerabilities, significantly outperforming a baseline that is based on tweet volume. Finally, we show how reports of severe vulnerabilities online are predictive of real-world exploits.

Multi-task Pairwise Neural Ranking for Hashtag Segmentation
Mounica Maddela, Wei Xu, Daniel Preoţiuc-Pietro
Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), pages 2538–2549
Abstract: Hashtags are often employed on social media and beyond to add metadata to a textual utterance with the goal of increasing discoverability, aiding search, or providing additional semantics. However, the semantic content of hashtags is not straightforward to infer as these represent ad-hoc conventions which frequently include multiple words joined together and can include abbreviations and unorthodox spellings. We build a dataset of 12,594 hashtags split into individual segments and propose a set of approaches for hashtag segmentation by framing it as a pairwise ranking problem between candidate segmentations.1 Our novel neural

approaches demonstrate 24.6% error reduction in hashtag segmentation accuracy compared to the current state-of-the-art method.

Neural Network Models for Paraphrase Identification, Semantic Textual Similarity, Natural Language Inference, and Question Answering
Wuwei Lan, Wei Xu
Proceedings of the 27th International Conference on Computational Linguistics (COLING), pages 3890–3902, **Best Paper Award**
Abstract: In this paper, we analyze several neural network designs (and their variations) for sentence pair modeling and compare their performance extensively across eight datasets, including paraphrase identification, semantic textual similarity, natural language inference, and question answering tasks. Although most of these models have claimed state-of-the-art performance, the original papers often reported on only one or two selected datasets. We provide a systematic study and show that (i) encoding contextual information by LSTM and inter-sentence interactions are critical, (ii) Tree-LSTM does not help as much as previously claimed but surprisingly improves performance on Twitter datasets, (iii) the Enhanced Sequential Inference Model is the best so far for larger datasets, while the Pairwise Word Interaction Model achieves the best performance when less data is available. We release our implementations as an open-source toolkit.

Character-Based Neural Networks for Sentence Pair Modeling
Wuwei Lan, Wei Xu
Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL), pages 157–163
Abstract: Sentence pair modeling is critical for many NLP tasks, such as paraphrase identification, semantic textual similarity, and natural language inference. Most state-of-the-art neural models for these tasks rely on pretrained word embedding and compose sentence-level semantics in varied ways; however, few works have attempted to verify whether we really need pretrained embeddings in these tasks. In this paper, we study how effective subword-level (character and character n-gram) representations are in sentence pair modeling. Though it is well-known that subword models are effective in tasks with single sentence input, including language modeling and machine translation, they have not been systematically studied in sentence pair modeling tasks where the semantic and string similarities between texts matter. Our experiments show that subword models without any pretrained word embedding can achieve new state-of-the-art results on two social media datasets and competitive results on news data for paraphrase identification.

4.3 Invited Talks

Leidos and team members were recognized for our expertise throughout the program as invited presenters at external venues.

Table 5: Invited talks for 2021

Presenter	Organization	Date	Topic/Venue
Wei Xu	Georgia Tech	Nov-21	Dataminr
Alan Ritter	Georgia Tech	Sep-21	University of Notre Dame NLP Seminar
Wei Xu	Georgia Tech	Sep-21	SimpleText Workshop at CLEF
Wei Xu	Georgia Tech	Aug-21	Stanford University (NLP Seminar)
Wei Xu	Georgia Tech	Jun-21	University of California, Los Angeles (Big Data and ML Seminar)
Kathleen Carley	Netanomics	Jun-21	"Socially Influence Campaigns: The Coordination of Events Using Bots and Misinformation," Plenary Panel, NSF Prpare workshop: Social, Behavioral, economic and governance aspects of pandemics – Virtual
Kathleen Carley	Netanomics	Jun-21	"State of Deepfakes and Disinformation Online," Deepfakes, Disinformation & Democracy (D3) Conference ,Opening Plenary Panel – Virtual
Kathleen Carley	Netanomics	May-21	"Disinformation and Manipulation during COVID-19 & the Election," Keynote, SAMSI – Virtual
Wei Xu	Georgia Tech	Apr-21	New York University (NLP and Text-as-Data Speaker Series)
Kathleen Carley	Netanomics	Jan-21	"The Power of High Dimensional Networks," Keynote, North American Social Network (NASN) – Virtual
Kathleen Carley	Netanomics	Jan-21	"Technology and Disinformation," MIT CSAIL IAP 2021: Workshop on Democracy – Virtual

Table 6: Invited Talks for 2020

Presenter	Organization	Date	Topic/Venue
Kathleen Carley	Netanomics	Dec-20	"Social Cybersecurity," Plenary, MORS Emerging Technology Forum (ETF) – Virtual
Wei Xu	Georgia Tech	Nov-20	Carnegie Mellon University (LTI Colloquium)
Wei Xu	Georgia Tech	Nov-20	University of Delaware, Newark, DE (ECE Department Seminar)
Kathleen Carley	Netanomics	Nov-20	"Social Cybersecurity: Observations and Directions," IDeaS Conference: Social Cybersecurity in Times of Crisis and Change, Carnegie Mellon University, Pittsburgh, PA – Virtual
Kathleen Carley	Netanomics	Oct-20	"Social Influence and Disinformation in Cyberspace," Open Science Symposium, Carnegie Mellon University, Pittsburgh, PA – Virtual
Kathleen Carley	Netanomics	Oct-20	"Disinformation and Manipulation in Cyberspace," Keynote – Greffenstette Symposium: Disinformation, Misinformation and Technology, Duquesne, Pittsburgh, PA – Virtual
Kathleen Carley	Netanomics	Oct-20	"Social Cybersecurity and the Election," Plenary, Disinformation One Month from the Election Virtual Round Table, CMU Computer Science
Kathleen Carley	Netanomics	Oct-20	Disinfodemic," Keynote – McGowan Symposium on Business Leadership and Ethics – Virtual
Kathleen Carley	Netanomics	Oct-20	"Disinformation Playground for Bots and Trolls," Keynote – WPFC – Virtual
Kathleen Carley	Netanomics	Oct-20	"Misinformation and disinformation in social media and the role bots play during the pandemic," Invited Plenary – SPB-BRiMS – Virtual
Kathleen Carley	Netanomics	Oct-20	"Social Cybersecurity," Keynote – ISAT/DARPA Digital Innoculation at Scale – Virtual
Wei Xu	Georgia Tech	Sep-20	Emory University, Atlanta, GA (CS Department Seminar)
Kathleen Carley	Netanomics	Jun-20	"Social Cybersecurity and the Pandemic." Keynote – ICWSM – Virtual
Wei Xu	Ohio State	Feb-20	University of Maryland, College Park, MD
Kathleen Carley	Netanomics	Feb-20	"Analysis of Social Networks for Intelligence," Plenary, Army War College Workshop, Carnegie Mellon University, Pittsburgh, PA.
Kathleen Carley	Netanomics	Feb-20	"Influence Campaigns in Social Media," Plenary, Hacking Democracy: Influence operations in the Digital Age, Oslo, NO
Wei Xu	Ohio State	Jan-20	University of Massachusetts, Amherst, MA

Table 7: Invited Talks for 2019

Presenter	Organization	Date	Topic/Venue
Alan Ritter	Ohio State	Dec-19	NeurIPS workshop on Conversational AI
Wei Xu	Ohio State	Dec-19	Georgia Institute of Technology, Atlanta, GA
Kathleen Carley	Netanomics	Dec-19	"Russian and Chinese Influence in Social Media," Plenary, Best Practices and Gaps to Be Closed—Addressing Foreign Interference in Democratic (and Other) Institutions, The Atlantic Council, Washington DC Cohosted by Carnegie Mellon University, Washington, DC
Kathleen Carley	Netanomics	Dec-19	"Advances and Opportunities in Social-cybersecurity," Plenary, DHS Foreign Interference, Washington DC
Kathleen Carley	Netanomics	Dec-19	"CUES - Understanding information maneuvers in social media using implicit information," Keynote - TextXD: Text Analysis across Domains, University of California, Berkeley, CA
Kathleen Carley	Netanomics	Nov-19	"Artificial Intelligence and Social Cybersecurity," Keynote - AAAI Symposium on AI in Government and Public Sector, Washington DC
Kathleen Carley	Netanomics	Oct-19	"Influence Operations: BEND maneuvers using bots and memes for social cybersecurity," Plenary, The NATO Science for Peace and Security Programme: Senior Leadership Roundtable on Information-Related Hybrid Threats in South East Europe, Ohrid, Macedonia
Kathleen Carley	Netanomics	Sep-19	"Social Cybersecurity," Plenary, Decadal Launch Event, National Academies, Washington DC
Kathleen Carley	Netanomics	Jul-19	"Social Cybersecurity," Plenary, DHS Conference on Disinformation
Alan Ritter	Ohio State	Jun-19	Twitter
Alan Ritter	Ohio State	Apr-19	University of Michigan
Alan Ritter	Ohio State	Apr-19	University of Maryland
Kathleen Carley	Netanomics	Apr-19	"Social Cybersecurity Dynamics," Keynote Computational Social Science – Quo Vadis? An Interdisciplinary Symposium Honoring Kathleen M. Carley, University of Zurich, Zurich Switzerland https://uzhdag.github.io/css/
Kathleen Carley	Netanomics	Apr-19	"Information Maneuver Assessment," Keynote - StratComAPAC2019, Singapore Singapore
Kathleen Carley	Netanomics	Apr-19	"Social Cybersecurity Dynamics," Keynote - SocialSens, Montreal Canada
Alan Ritter	Ohio State	Mar-19	Rice University
Alan Ritter	Ohio State	Jan-19	University of Wisconsin

Table 8: Invited Talks for 2017 and 2018

Presenter	Organization	Date	Topic/Venue
Kathleen Carley	Netanomics	Nov-18	"Social Cybersecurity," Keynote - IEEE Intelligence and Security Informatics, Florida International University, Miami, FL
Kathleen Carley	Netanomics	Oct-18	"Social Cybersecurity," Keynote MOBICOM, Delhi, India
Kathleen Carley	Netanomics	Sep-18	"Network Analysis and AI for Humanitarian Assistance and Disaster Response," Plenary, AI and HADR workshop, AI Visionary Talk, Carnegie Mellon University, Pittsburgh, PA
Wei Xu	Ohio State	Jun-18	Midwest Machine Learning Symposium, Chicago, IL
Alan Ritter	Ohio State	May-18	Stanford University
Wei Xu	Ohio State	May-18	Facebook, Menlo Park, CA
Wei Xu	Ohio State	May-18	Twitter, San Francisco, CA
Wei Xu	Ohio State	May-18	Stanford Research Institute, Menlo Park, CA
Alan Ritter	Ohio State	Dec-17	NeurIPS workshop on Learning with Limited Labeled Data
Wei Xu	Ohio State	Nov-17	IBM Thomas J. Watson Research Center, New York
Kathleen Carley	Netanomics	Nov-18	"Social Cybersecurity," Keynote - IEEE Intelligence and Security Informatics, Florida International University, Miami, FL

5 Estimates of Feasibility

The Leidos team was able to successfully execute the program plan, resulting in significant advances in the scientific understanding of online information spread related to the overall program goals. The PEGASIS solution to identifying, collecting, processing, enriching, annotating, anonymizing, and delivering research quality corpora is robust and reproducible. However, at least one of the principal challenges that Leidos identified at program outset will always remain: the online information environment is constantly evolving, and systems designed to exploit it must continuously evolve as well.

5.1 Lessons Learned

5.1.1 Data Collection

Collecting research-quality corpora of real, user-generated social media activity presents many problems. Notably, each platform provider maintains a unique set of terms of use that must be adhered to, both from a legal and ethical perspective. Oftentimes over the course of a four-year program, these terms change, which require a dynamic collection strategy. At other times,

these terms effectively prohibit the use of bulk volumes of data from a desired platform (i.e. Facebook, Weibo, etc.) without invoking additional USG authorities.
Second, API endpoints and data schema evolve over the life of a program like Social Sim. At times, platforms will enable additional features to be extracted via a webUI than via the bulk API. As the availability of these organic data features evolves, the collection team must adapt accordingly, including developing ad hoc collection mechanisms, rebuilding data features from multiple collection mechanisms, or communicate with the research teams why certain features are not available at scale.

5.1.2 Enrichment

Several highly desirable publicly available enrichments are made available via API by independent third parties using a tiered pricing plan. With a free but rate-limited option available, researchers often have experience working with and ask for a certain enrichment. Unfortunately, scaling this enrichment to the millions of events or users in our corpora requires either: A) paying for premium access, B) creating multiple accounts to simultaneously access the API, and/or C) planning for a multi-week enrichment effort. Often all three are necessary. Like the primary data sets, the availability of publicly available enrichments evolves over the life of the program. Some services maintained by third parties are unexpectedly terminated, rate limits are imposed, or the features available change with upgrades. Being dynamic and adaptive to this reality is a must.

5.1.3 Timing

TA1 teams desire training data as early in the process as possible. Proper data exploration, collection, processing, enrichment, and preparation takes several months for a single Challenge Problem. In most cases, due to program rhythms, we were finalizing the selection of Challenge Problems *after* the completion of the previous CP, which created a necessary multi-month lag before new data could be delivered to the TA1 teams for the next CP. In some cases, this meant delivering training data to the TA1 teams with limited time for experimentation and training. An improved program execution model would stagger data collection and preparation a full CP cycle ahead of the TA1 research challenge. This might require giving the data provisioning team a head-start but would maximize research teams' time with data.

Annex A: CP4 Venezuela Narrative Annotation Guidelines

Annotators will read tweets and youtube texts (comments, titles, etc.) and annotate them for **stance** and **narrative**. Annotators should make every attempt to make annotations based on their understanding of the original language, but should also utilize google translate as necessary to affirm their translation. When the text includes information for which contextual knowledge is assumed (i.e. a named person or event, a concept like "**cacerolazo**", etc.), or the annotator is otherwise unsure of the meaning of a term, the annotator should perform quick internet research to inform their annotation. Annotators SHOULD NOT attempt to review the tweet or comment from the original social media web portal. When annotating for narratives, annotators should enter the narrative labels as comma separated strings, and they should only enter narrative labels that occur in this document.

We are using a 8 to 1 ratio of single-annotator annotations to all-annotator annotations. This means that for every 8 documents that only one of us annotates (split equally), there is one document that we all annotate. We will use Cohen's kappa for inter-annotator agreement.

Stance:

A stance is a point of view on a topic. These points of view should be very high level and should represent a user's attitude (usually for or against) a **topic**. A topic in this context should be understood as the central focus of a dataset, which is strongly influenced by the collection query. A topic may be an event, entity, or concept. The topic of interest will be consistent across a dataset.

In the Venezuela dataset, the topic is twitter outages which occurred in the context of the Venezuelan presidential crisis. Because of the context, most posts fall on one of two sides in the Venezuelan presidential crisis: a group that favors the Maduro government and is hostile toward Juan Guaido and the US, and a group that favors Guaido and disapproves of Nicolas Maduro or his administration. There are also posts that are neutral or unclear in their stance. In the Venezuela data, posts should be considered part of the anti-maduro stance group if they **express disapproval of Nicolas Maduro, his administration, or their actions. Posts which express approval of Guaido, his administration, or their actions should also be considered part of the anti-Maduro stance group.** The reverse is true for the pro-Maduro stance group. Posts which do not take a stance in the crisis or which are unclear should be put in a third group.

Narratives:

In this context, narratives are recurring statements which are used to express a point of view on a particular topic. Narratives may be explanations of events, interpretations of the motives of actors, statements which emphasize specific concepts, or other techniques to express a point of view. While stance is the point of view itself, a narrative is a particular idea or claim which supports the stance.

Stance Groups:

Anti-Maduro: documents should be labeled with the stance am (anti-maduro) if they either:

- Show negative opinion of Maduro or his administration
- Show positive opinion of Guaido (his primary opponent for presidential legitimacy)
- Promote actions that are in opposition to decisions by the Maduro government (e.g. accepting humanitarian aid from sources that have been rejected by the Maduro government)
- Use hashtags or phrases that are clearly aligned with the opposition to Maduro, unless it is clear they are used sarcastically or otherwise inauthentically
    - Some hashtags include: VenezuelaGritaLibertad, VenezuelaLibre, 23ECalleHastaQueSeVaya, GritemosConBrio, Usurpador, GuaidóPresidente
    - Hashtags which **do not necessarily** entail am stance, but often do: 23e, 23f, 23ene, 23feb, VamosBien, Guaido, VenezuelaAidLive

Pro-Maduro: they should be labeled pm (pro-maduro) if they:

- Show positive opinion of Maduro or his government (including the Venezuelan military)
- Show negative opinion of Guaido
- Show positive opinion of Chavez or Chavismo
- Use hashtags or phrases that are clearly aligned with Maduro or Chavismo
    - Some hashtags include: LealesSiempreTraidoresNunca, WeAreMaduro, YankeeGoHome, HandsOffVenezuela, LasCallesSonDelChavismo, VenezuelaYElMundoConMaduro, UnidadYLealtadConNicolas, LaHoraDeLosCLAP, VenezuelaEnDefensaDeLaPaz

Unclear or Neutral: they should be labeled with a question mark "?" if they:

- Do not appear to take a side (e.g. describe events in a way that doesn't seem biased)
- Favor some other actor that is opposed to both Maduro and Guaido (this should be rare based on our experience)
- Lack information, context, or are unclear. This includes documents which do not clearly mention a relevant party or hashtag.

Narrative Groups:

We identified 10 narrative groups: Guaido, assembly, Maduro, protests, arrests, violence, international, military, crisis, and other. These groups were established by consensus among three annotators after initial exploration of the data and a first attempt at narrative labeling. We attempted to make larger scale semantic groups that correspond to important recurring aspects of narratives, based on our best judgment as to what is relevant after reading example documents. We have also used common disagreements among annotators as feedback to modify the narrative groupings. While some of the top-level narratives express important information on their own ("protests"), others have no well-defined meaning or are too vague

on their own ("other"). This means that of the 10 top-level groups, **5 can occur on their own as a narrative: protests, arrests, violence, military, and crisis.** There are 49 total sub-narratives.

Guaidó: these narratives focused on Juan Guaidó or his administration. “Guaido” should not be used alone without a sub-label. Documents that only express general support or disapproval of Guaidó should be labeled only with the stance “am” or “pm” respectively.

- **Guaido/legitimate, Guaido/illegitimate, Guaido/legitimate/international, Guaido/illegitimate/international:** documents that state or imply that Guaido is legitimate or illegitimate as the president of Venezuela should be marked with the label “guaido/legitimate” or “guaido/illegitimate”. Further, if they express his legitimacy from the perspective of international support and recognition (from representatives of foreign governments including members of congress/parliament, foreign ministry, etc.), they should be marked with “guaido/legitimate/international”. If they merely state that he is the president without further clarification then they should only have the label “guaido/legitimate”. Example:
    - “Se sigue avanzando en la Unión Europea para el reconocimiento y apoyo pleno de nuestra [refers to Guaido and supporters] lucha legítima y constitucional. Agradecemos las palabras y el compromiso adquirido por el Presidente del Gobierno de España, Pedro Sánchez @sanchezcastejon #Venezuela #26Ene.” (guaido/legitimate/international)
- **Guaido/US_support:** Documents that state that the US supports Guaido should be marked with guaido/us_support. This should include US officials. This narrative can be used in an anti-Maduro way that shows that Guaido has international legitimacy, and in that case guaido/legitimate/international should also be applied. However, the narrative could also be used in a pro-Maduro way if it emphasizes that Guaido is a US “puppet” or similar statements negatively linking Guaido to the US. In these cases guaido/legitimate/international should not be used since the author doesn't view Guaido as legitimate. Example:
    - “Trump just declared an opposition leader, who was not elected by the people, the “Interim President" of #Venezuela. President #Maduro then officially cut off diplomatic ties with the US government, giving American diplomats 72 hours to leave. They've refused. #HandsOffVenezuela”

Assembly: Documents that emphasize the legitimacy of the Venezuelan National Assembly should receive the label ”assembly/legitimate”. Documents that express that the assembly is not legitimate should receive “assembly/illegitimate”. Example:

- “Nicolás no le garantiza nada a nadie. La única ruta para sacar a Venezuela adelante es la planteada desde nuestra @AsambleaVE. Nuestra ruta es muy clara: cese de la usurpación, gobierno de transición y elecciones libres. El llamado es a la gente a que nos acompañe este #23Ene.” (assembly/legitimate)

Maduro: narratives about Nicolas Maduro or his administration. The label "maduro" should not be used alone.

- **Maduro/legitimate, Maduro/illegitimate, Maduro/legitimate/international, Maduro/illegitimate/international**: these narratives follow the same rules as the Guaidó legitimacy narratives, but instead refer to Maduro. Maduro is often referred to as a "usurper" without direct mention of his name, and this can be inferred to mean that he is illegitimate. When it can safely be assumed that the person being discussed is Maduro, the label "maduro/illegitimate" should be applied. If Maduro is referred to as "the president" or "president Maduro", it can be inferred that he is legitimized. Example:
    - "#AHORA La posición unánime en este encuentro entre los embajadores @JulioBorges, @carlosvecchio, Gustavo Tarre y @Almagro_OEA2015 es que no se le pueden pedir elecciones a Nicolás Maduro porque de plano saben que sus resultados (con este CNE) serán fraudulentos. #30ene" (maduro/illegitimate)
- **Maduro/narco**: documents that state a connection between Maduro and narco-trafficking, or that call him a "narco-dictator", should have the label "maduro/narco". Example:
    - "Hace 61 años #Venezuela derroco una dictadura. Hoy, como aquel #23E de 1958, es un día histórico para los venezolanos quienes marchan en contra de la narco-dictadura de #Maduro y alzan su voz en apoyo al Presidente Provisional de #Venezuela @jguaido."
- **Maduro/dictator**: this label should be applied when Maduro is described as a "dictator" or "tyrant", when Maduro is accused of "oppression", or otherwise when the Maduro administration is represented as an oppressive dictatorship. Example:
    - "Condenamos la represión del usurpador @NicolasMaduro que ahora está pasando en las calles de Caracas contra el pueblo de #Venezuela que hoy reclama por su libertad y contra la tiranía. #23Ene #OEAconVzla"
- **Maduro/events, Maduro/events/pro, Maduro/events/anti**: applied when the document describes events such as marches or concerts organized by or in support of Maduro or his administration. When the document expresses support for the event, the label "maduro/events/pro" should be used, but when it expresses disapproval of the event (e.g. states that there was low turnout), the label "maduro/events/anti" should be used. Example:
    - "So far the #Maduro concert not looking too good." (maduro/events/anti)
- **Maduro/cuba_support:** States or implies that the government of Cuba provides overt or covert support to the Maduro government. Example:
    - "DEVELOPING: We are receiving intercepted transmissions of Cuban agents present & directing repressive actions in Ureña. #23Feb"
- **Maduro/russia_support:** States or implies that the government of Russia provides overt or covert support to the Maduro government. Example:
    - "Al menos 400 paramilitares rusos habrían viajado a Venezuela, para reforzar la seguridad y así proteger a Nicolás Maduro, frente a las protestas antigubernamentales que se ha realizado en el país #27Ene"

Protests: The label “protests” should be used when the document describes anti-maduro protests, rallies, or other public demonstrations not including the Venezuela aid concert. This can also be applied when the document describes an event but it’s not clear who the event supports. Events that can be inferred to support the Maduro government should receive the label “maduro/events” or its sub-labels instead. If the document describes arrests or violence against attendees of anti-Maduro events use “arrests/opposition/protesters” or “violence/against_opposition/protesters” instead. Violence by protesters against police or military should receive the labels “protests” and “violence/against_maduro”. Violence at a protest that has an unclear source or target (e.g. “shots heard near a protest”) should receive the labels “protests” and “violence”. Example of “protests” tag:

- “#AHORA Habitantes de San Martín, protestan en diferentes puntos, entre ellos la entrada de El Guarataro, donde iniciaron una quema de basura y cauchos en manera de rechazo a la usurpación del poder por parte de Nicolás Maduro. #23Ene”

Arrests: describes arrests or people who have been taken prisoner. The label “arrests” should be used alone when the document describes arrests with no further information, or if the person arrested does not fall into one of the following groups.

- **Arrests/opposition:** arrests of officials opposed to Maduro, or of people who can be inferred to be politically opposed to Maduro. The phrase "political prisoners" can be inferred to denote the arrest of opposition. Example:
    - “Por Óscar Pérez y su grupo, por los militares que están presos por querer detener la carnicería chavista, por todos los presos políticos, por Franklin Brito. #VenezuelaLibre #2Feb”
- **Arrests/opposition/media:** arrests of journalists or media affiliates. Example:
    - “¡Urgente! Por orden del despacho de la presidencia de la republica, ordenan diferir por cuarta vez la audiencia del reportero Jesus Medina Ezaine, quién está secuestrado en la carcel militar de Ramo Verde. Llamó al país a que no tengan miedo y luchen viva Venezuela libre carajo..”
- **Arrests/opposition/protesters:** arrests of protesters or attendees of anti-Maduro events. If this tag is used, adding “protests” is not necessary since it can be inferred. Example:
    - “#31Ene 1:40PM En este momento 19 menores son excarcelados en Yaracuy que fueran detenidos el #23Ene por protestas Gracias y Felicitaciones equipo @ForoPenal Yaracuy @GabrielGalloG En poco será excarcelada Laura Gallo, madre de nuestro coordinador @ForoPenal En poco colocamos foto”

Violence: narratives that describe acts of violence or intimidation such as threats. Violence should be interpreted as physical harm in addition to threats or acts clearly meant as threats such as firing weapons into the air. The label “violence” should be used alone when the description is lacking information, or it does not fit into one of the following categories.

- **Violence/against_opposition:** violent acts committed against people who are politically opposed to the Maduro government. Example:
    - "#AHORA Acaban de golpear a José Manuel Olivares en vivo, en transmisión de Univision @UniNoticias #23F"
- **Violence/against_opposition/protesters:** violence against protesters or attendees of anti-Maduro events. If this tag is used, adding "protests" is not necessary since it can be inferred. Example:
    - "#23Ene Siete (7) personas fallecidas en contexto de protestas el día de hoy: Táchira 2, Bolívar 1, Portuguesa 2 , Amazonas 2. Otros 2 fallecidos en Bolívar, por verificar. #Venezuela @ForoPenal"
- **Violence/against_maduro:** violence against people who are politically aligned with Maduro or his administration (including the military). Example:
    - "grupo de opositores drogados en San Martin Caracas intentan voltear camión de la GNB PARA quemar lo @NicolasMaduro @VenezuelaH4ck @lubrio @Mamba4F92 @OrlenysOV @RalitoDigital @PatriciaDorta40 @GuerrillaZK @VTVcanal8 #WeAreMADURO #YankeeGoHome"

International: narratives about international involvement in the events in Venezuela. The label "international" should not be used alone as it doesn't convey much information.

- **International/aid:** Descriptions of foreign humanitarian sent or requested to be sent to Venezuela. Discussion of the Venezuela Aid Live concert with no other information should not be given this label unless it is clear that the main concern of the author is humanitarian aid. This is because it's common for people to discuss the musicians or music at the aid concert only, and it's not clear whether people are primarily concerned with the aid or with the music itself. Example:
    - "#24Ene Informa Marco Rubio que @jguaido ha solicitado formalmente la ayuda de los Estados Unidos para colaborar y proporcionar al pueblo de #Venezuela alivio humanitario inmediato."
- **International/aid_rejected:** describes rejection or destruction of foreign humanitarian aid by the Maduro government.
    - "#AHORA Incendian tres gandola de #AyudaHumanitaria en puente internacional Francisco de Paula Santander en #Ureña. Testigos aseguran fueron policías bolivarianos junto a colectivos. Personas trata de salvar parte de la carga #23Feb @ReporteYa"
- **International/emigration:** describes emigration from Venezuela or references Venezuelan expatriates. Example:
    - "Agradecido con los venezolanos en #Miami por su expresión masiva de solidaridad con la #AyudaHumanitaria para detener el dolor de tantas familias que hoy sufren en #Venezuela ¡Así somos los venezolanos! Uds. han sido clave en todo este esfuerzo humanitario para este #23F"
- I**nternational/respect_sovereignty:** emphasizes Venezuela's national sovereignty, independence, or right to conduct its own political processes. Documents that describe US intervention, imperialism, or that generally state the US should not be involved in

Venezuelan affairs should also receive this label. Documents containing #handsoffvenezuela are almost always in this category. These themes may be coded in language such as "defending the homeland" against "the empire" (the US). Example:
    - "#LealesSiempreTraidoresNunca no es la primera vez que algún imperio ha querido apoderarse de nuestra Patria, los hemos derrotado y los hemos echado del suelo patrio, hoy con más razón los derrotaremos, en unión Cívico Militar Nosotros Venceremos"
- **International/US_sanctions:** mentions us sanctions against Venezuela. Example:
    - "#12Feb Reuters: Maduro busca ayuda de la OPEP contra sanciones de EEUU. La OPEP dice que le preocupan las normas petroleras, no la política. Rick Scott: La OPEP no prestará ayuda a Nicolás Maduro."
- **International/break_us_relations:** mentions that the Maduro government broke diplomatic relations with the US. Example:
    - "U.S. provocatively refuses to withdraw its personal in 72 hours as commanded by Venezuelan leader Maduro. After 72 hours, from the Maduro state perspective, US will be in the country illegally, effectively a small occupation, leading to additional confrontations."
- **International/military:** discussion of military buildups or operations by countries foreign to Venezuela. This also includes references to a "military intervention" in venezuela. Example:
    - "#AHORA El ejercito Brasileño en camino a la Frontera Venezolana #santaelenadeuairen #23f #23FAvalanchaHumanitaria"

Military: narratives about the Venezuelan military, security forces, or other armed militarized organizations that report to the Maduro government. Documents that mention the military and have no further information or do not fit into one of the following categories should receive the label "military" alone. This would include mentions of military personnel or branches. Some common phrases that denote what we consider militarized groups are: FANB (national Bolivarian armed forces), DGCIM (Directorate General of Military Counterintelligence), GNB (Bolivarian national guard), GN (national guard), FA or FFAA (armed forces), PNB (national Bolivarian police), FAES (police special actions force), SEBIN (Bolivarian national intelligence service), "colectivos" (armed paramilitary groups reporting to the Maduro government).

- **Military/desertions:** documents that describe or call for desertions or defections of military members. Example:
    - "#23FAyudaHumanitaria is here. 3 members of National Guard of #Venezuela deserted moments ago at Simón Bolívar International Bridge according to official information. Remember these guard members are eating very little & have had high rates of desertion over the last 18 months."

Crisis: narratives that refer to the Venezuelan humanitarian crisis. Documents that only mention a "humanitarian crisis" or "crisis" should have the label "crisis" alone.

- **Crisis/lack_essentials:** states that there is a shortage of food or medicine in Venezuela, or gives examples of affected people. Discussion of humanitarian aid alone does not fall under this label. Example:
    - “Amigos Miguel aún tiene un cuadro de desnutrición muy severa y necesita alimentos proteicos, está recibiendo nutrición Parenteral. Presidente @jguaido Miguel necesita ser uno de los primeros venezolanos en recibir la ayuda humanitaria que entrará el #23F #Venezuela👉 @GotasEVzla”
- **Crisis/looting:** describes riots or looting. Example:
    - “Disturbios frente a la casa “Robert Serra” en el sector conocido como Puerta Caracas, la noche de este lunes #21Ene / video @ReynthardtM”

Other: narratives which don’t fit into the above groups.

- **Other/chavez, other/chavez/pro, other/chavez/anti:** uses terms related to Chavez and his legacy like “chavez”, “chavismo”, “chavista”. Documents that discuss Chavez negatively should have the label “other/chavez/anti”, and those that discuss him positively should have the label “other/chavez/pro”. Example:
    - “Hoy #23Feb el pueblo se moviliza en Caracas y en todas las ciudades del país. Vamos todas y todos a las calles a defender nuestra independencia, con conciencia y alegría. No habrá guerra en la Patria de Bolívar y Chávez, aquí triunfará la paz. ¡Venezuela se Respeta!” (other/chavez/pro)
- **Other/censorship_outage**: describes outages of media platforms (including social media and traditional media) in terms that clearly imply censorship.
    - “#AlertaSNTP | Supercable se une a las cableoperadoras que sacan del aire el #VenezuelaAidLive, en la frontera, ya lo habían hecho DirecTv e Intercable. La Censura no borrará este día en que lo mejor de la música latina se unión por el rescate de la libertad en Venezuela #22Feb”
- **Other/anti_socialism:** identifies socialism as the cause of the Venezuelan crisis or otherwise disparages socialism, communism, or ”leftism”. Example:
    - “Chavista: Ayer #29dic Mun Paez | Pernil podrido te manda tu Pdte #Maduro | Te aplicaron el manual cubano: racionamiento para controlarte y tu pendejo mendigando limósnas al régimen socialista | Comes lo que te mereces y eso es lo que te da tu gobierno”
- **Other/anti_capitalism:** disparages capitalism. Example:
    - “This man bought an entire island without the consent of Virgin Islanders, kicked out the natives, and charges people thousands of dollars a night to visit. If Virgin Islanders go back to their island now owned by capitalist pig Richard Branson, they are arrested.”
- **Other/restore_democracy:** calls for increased democratic processes in Venezuela. Example:
    - “Nuestras felicitaciones a @jguaido como Presidente encargado de #Venezuela. Tiene todo nuestro reconocimiento para impulsar el retorno del país a la democracia #23Ene #OEAconVzla”
- **Other/request_observers:** calls for greater media coverage of events in Venezuela.

    - "#Venezuela este es uno de los #vídeos que la Narco-dictadura de Maduro no quiere que el #mundo vea .. Periodistas ingresaron al Hospital Universitario de #Caracas capital venezolana.. ahora imaginen cómo deben estar los demás hospitales y centros sanitarios."
- **Other/planned_coup:** refers to the events of the presidential crisis as a "coup" with the intent to remove Maduro and install Guaido. Documents assigned to this narrative may or may not specifically mention Guaido. Example:
    - "A note from Roger: THIS IS TODAY!!!!!!!!!!!!!!!!!!!! STOP THIS LATEST USG INSANITY, LEAVE THE VENEZUELAN PEOPLE ALONE. THEY HAVE A REAL DEMOCRACY, STOP TRYING TO DESTROY IT SO THE 1% CAN PLUNDER THEIR OIL. US HANDS OFF #VENEZUELA! #NICOLASMADURO #STOPTRUMPSCOUPINVENEZUELA"
- **Other/media_bias:** states that the international media, "western" media, or specific media outlets including Venezuelan media do not fairly portray the events in Venezuela. Example:
    - "[VIDEO] Así estuvo el estado Falcón este #30Ene. Fue la marcha antiimperialista y contra el intervencionismo, conducida por @dcabellor y el gobernador Víctor Clark. A pesar de la cantidad de gente, fue un evento censurado por la mayoría de los medios acreditados en #Venezuela."

Annex B: CP5 Belt and Road in Pakistan Narrative Annotation Guidelines

1 Introduction 1.1 Annotation

We will annotate Twitter and YouTube texts for stance, narrative protagonist/antagonist, and framing. The annotation set will be randomly selected from all unique text values, weighted by the number of retweets or duplicate texts, in order to direct the annotation toward texts that will cover the largest portion of the dataset. The annotation set will be divided into equal sections for each annotator and an additional equally sized section that all annotators will annotate. After annotation, Cohen's Kappa agreement scores will be computed for each label and overall.

2 Stance

Based on the 2016 SemEval Twitter stance detection task, we understand stance detection as "automatically deter- mining from text whether the author of the text is in favor of, against, or neutral towards a proposition or target" [2]. In this scenario, the target of the stance is the CPEC (China-Pakistan Economic Corridor). There will be three relevant stance values: pro-cpec, anti-cpec, and neutral/unknown. The three stance values are described below:

Pro-CPEC Shows a positive attitude toward CPEC, individual CPEC projects, or toward cooperation between China and Pakistan in general.

Anti-CPEC Shows a negative attitude toward CPEC, CPEC projects, or countries involved (China or Pak- istan).

Neutral/Unknown The stance of the author toward CPEC is either neutral or it is not clear from the text due to lack of information.

3 Narrative Protagonist/Antagonist

In addition to stance, we will annotate for two more fields: protagonist and antagonist. A protagonist is understood as the main character of a narrative and, although a reader does not universally have a positive attitude toward a protagonist, they often have the sympathy of the reader. An antagonist is in conflict with the protagonist, so more often than not the reader may have a negative attitude toward the antagonist. We use the terms protagonist and antagonist to refer to a specific set of relevant entities when they are the target of positive or negative sentiment in a document. The set of relevant entities includes four state actors relevant to the scenario: Pakistan, China, India, and the USA. A document may have more than one protagonist or antagonist, or it may lack a protagonist or antagonist. In order to count as a protagonist or antagonist, the actor must be involved in an action as either subject or object, rather than just mentioned in a text. State actors should be marked as a protagonist or antagonist if they are specifically mentioned (e.g. the text uses "India" or "Indian").

Example: (Auto-translated Urdu) This is the network which in a very systematic manner launched a false campaign against #CPEC, Pakistan Army and Lieutenant General Asim Bajwa, created a fake website and broad- cast content on it without any evidence. The entire operation was commanded by Major Guru Aryan of the Indian Army.

Protagonist: Pakistan Antagonist: India

4 Frames

We will annotate for frames in documents, using a definition of framing similar to the Media Frames Corpus [1]:

“To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described.”

We have identified 21 frames grouped into a hierarchy based on both shared topic (e.g. the “controversies/china” frame group contains frames for political controversies related to China), and shared stances (groups should mostly contain frames that lean pro-cpec or anti-cpec). The hierarchy of the frames is shown in figure 1.

4.1 Frame Group: Controversies

The frames in the “controversies” group refer to political controversies related to either Pakistan or China.

4.1.1 Frame Group: Controversies/China

The frame group “controversies/china” contains frames about political controversies regarding China.

4.1.1.1 Frame: Controversies/China/Exploitation The frame “controversies/china/exploitation” should be used to tag documents that characterize CPEC or the Belt and Road Initiative as exploitation or colonialism, or emphasize Chinese mistreatment of the local population in relevant countries. It may also describe loaning practices that are perceived as predatory.

Example: CPEC was supposed to make Pakistan a super power...a land of immense wealth & prosperity...a country that India would be jealous of.

One day they realised that China had scammed them and stolen their money, pride & sovereignty. This is their story...

4.1.1.2 Frame: Controversies/China/Funding Documents that mention that China or Chinese financial institutions are having difficulty funding CPEC projects should have the frame "controversies/china/funding"

Example: CPEC has reportedly hit snags as the Chinese banks/ financial institutions are reluctant to fund the projects due to current state of affairs This financing issue was raised by Shah Mehmood Qureshi, during his visit to China last week.

4.1.1.3 Frame: Controversies/China/Border The frame "controversies/china/border" applies to documents that negatively discuss China's involvement in border disputes with India.

Example: For the the first time since Kargil, Indian territory occupied by foreign soldiers 5,000 Chinese soldiers are in Galwan/Pangong
Nepal revives a border dispute
Terrorism in Kashmir has increased

China-Pak-Nepal are working together. What is the Modi Govt's response?

4.1.1.4 Frame: Controversies/China/Uighur The frame "controversies/china/uighur" should be applied when the document brings attention to the Uighur controversies in Xinjiang China.

Example: #justiceforUighurMuslims #china #CPEC #PMImranKhan

4.1.2 Frame Group: Controversies/Pakistan

This frame group includes frames that refer to political controversies related to Pakistan or Pakistani officials. 4.1.2.1 Frame: Controversies/Pakistan/Army Negatively discusses the Pakistani army or its conduct.

Example: Balochistan is the most deprived region in Pakistan
CPEC is behind the killings, disappearances and human rights violations in Balochistan Many villages, which were in the CPEC route, are being razed by Pak Army
Why are Human Rights activists mum? #ChinaPakExploitationCorridor

4.1.2.2 Frame: Controversies/Pakistan/Bajwa Discusses reported financial wrongdoing or corruption by Asim Bajwa.

Example: Damning indictment of former DGISPR & CPEC Auth. head Gen Asim Bajwa: The Bajwa family's companies spent an estimated $52.2 million to develop their businesses & $14.5 million to purchase properties in the Asim Bajwa's sons joined the Bajco Group companies in 2015 #BajwaLeaks

4.1.2.3 Frame: Controversies/Pakistan/Baloch Mentions Baloch nationalism, independence, or related in- dividuals in a positive light, or characterizes Pakistan or China as mistreating Balochistan or Baloch people.

Example: Even before #ArifWazir was buried, Pakistan military killed two Baloch young men in Kalat, #Shah- dadBaloch & #EhsanBaloch, graduates of Quaid-e-Azam University. Both were critical of the Pakistani establishment for exploitation of Balochistan's natural resources & rights violation

4.1.2.4 Frame: Controversies/Pakistan/PTI Negative statements about Imran Khan or the PTI (Pakistan Tehreek-e-Insaf) party. This may come from outside Pakistan or it may come from within Pakistan among people expressing support for the PMLN party or Nawaz Sharif.

Example: Public Debt 2013 = Rs 14 tr 2018 = Rs 24 tr Mar, 2020 = Rs 35 tr (Source Budget2020-21 doc) #PTI borrowed as much as #PMLN borrowed in 5 yrs within 21 months. PMLN installed 11,000 mw energy, built 2,000kms mways, CPEC. What did IK do with loans?

4.1.2.5 Frame: Controversies/Pakistan/Students Negative statements about a controversy regarding Pak- istani students who were unable to return to school in China.

Example: Do you care about students future? No? Not a single statement from Pakistan embassy regarding stu- dents studying in china stuck in Pakistan! #PakistaniStudentsWantBackToChina #TakeUsBackToSchool #TakeUsBackToChina

4.2 Frame: Ineffective

States that CPEC projects are ineffective, mismanaged, wasteful, or otherwise fail to produce the intended benefits.

Example: Punjabi Army & Chinese Communists' joint venture.. The world famous CPEC ... could not sustain the rain & flooding water. I used to hear that there are heavy punishments for such defective constructions in China. CPEC & Pakistan are on self destruction mode.

4.3 Frame Group: Opposition

This frame group contains frames that are largely supportive of CPEC but are negatively discussing opposition to CPEC.

4.3.1 Frame: Opposition/Propaganda

Discusses negative coverage of CPEC and characterizes it as propaganda, false, or misinformed.

Example: (Auto-translated Urdu) My take on the FALSE propaganda against Retd. Lt. Gen. SAsimSBajwa by PMLN Media Cell & Indian Trolls

The rapid success of Gwadar Port and CPEC and the development of Balochistan have eroded billions of rupees of Indian propaganda. They cry.

Frame: Opposition/Kashmir

4.3.2

Has a negative attitude toward the Indian occupation of Kashmir or states Kashmir should be part of Pakistan

Example: Love you Pakistan from Indian illegally occupied Kashmir

4.4 Frame Group: Leadership

Frames that positively discuss Pakistani leaders or officials.

4.4.1 Frame: Leadership/Bajwa

Discusses Asim Bajwa, retired general and chairman of CPEC, in a positive way.

Example: The way Asim Saleem Bajwa has breathed new life into #CPEC over the past few months, Asim Saleem Bajwa Pakistan for India. . .

4.4.2 Frame: Leadership/Khan

Positively mentions Imran Khan, prime minister of Pakistan.

Example: #Pakistan ’s successful Foreign Policy in ONE picrure!
•CPEC growing fast
• US acknowledging role in US-Afghan Taliban Peace Talks
• Saudia REFUSING to accept Israel after #ImranKhan refuses to accept Israel •Turkey, Malaysia valuing more
WELL DONE, Imran Khan!

4.4.3 Frame: Leadership/Sharif

Positively discusses leadership by Nawaz Sharif, former prime minister of Pakistan. Some users may bring attention to Sharif to contrast with Imran Khan.

Example: (Auto-translated Urdu) Mian Nawaz Sharif endured cruel punishments for the sake of this country. His crime was that he came first in the service of this country. Be it nuclear blast or CPEC first thought of Pakistan #salam nawazsharif

4.5 Frame Group: Benefits

Group containing frames related to benefits to Pakistan as a result of CPEC or cooperation with China.

4.5.1 Frame: Benefits/COVID

Discusses Chinese assistance to Pakistan to help with COVID-19.

Example: China has provided emergency assistance to Pakistan for #COVID19 prevention and control, including 12,000 test kits, 300,000 masks, 10,000 protective suits and 4 million U.S. dollars to build hospitals, according to ambassador Yao Jing CathayPak

4.5.2 Frame Group: Benefits/Connections

These frames discuss new connections between Pakistan or CPEC and outside countries in a positive way.

4.5.2.1 Frame: Benefits/Connections/Afghanistan Positive discussion of connections between Pakistan or CPEC and Afghanistan in terms of trade or infrastructure.

Example: Pakistan's decisions to open routes for Afghan trade are "positive" and the #CPEC under the China- proposed #BRI can greatly enhance regional connectivity and "link Pakistan with Central Asia via Afghanistan," says Pakistani official

4.5.2.2 Frame: Benefits/Connections/Iran Positive discussion of connections between Pakistan or CPEC and Iran in terms of trade or infrastructure.

Example: #India lost big time by opposing $62bil #CPEC but will lose even bigger with #ICEP $400bil 25year pact as part of #BRI "#Iran #China Economic Pact"

It'll result in #CPEC extending to Iran, #Turkey #Russia & #India getting kicked out of Chahbahar.

4.5.3 Frame Group: Benefits/Development

Group containing frames about development of infrastructure in Pakistan as a result of CPEC.

4.5.3.1 Frame: Benefits/Development/Energy Positively discusses energy or power plant projects in Pak- istan related to CPEC.

Example: PM AJK & CEOs of China Three Gorges & China Gezhouba visited separately to thank CPEC Authority for helping realise Kohala Power & Azad Pattan power Projects & further coord for their execution. $4 Bn Investment, 1800 MW cheaper Hydel power with 8000 jobs. #cpec #CPECMakingProgress

4.5.3.2 Frame: Benefits/Development/Maritime Mentions maritime or port development projects such as Gwadar in a positive way.

Example: Building South Balochistan: Tender for important section of M-8 construction. Will improve connectiv- ity of Gwadar Port,revolutionise socio-economic dev of the region, bring prosperity, address long term deprivation ,Something people been waiting for decades #cpec #CPECMakingProgress

4.5.3.3 Frame: Benefits/Development/Misc Miscellaneous development projects such as airports or fiber optic cable installation. This also includes housing construction and education or training initiatives.

Example: Construction of Gwadar International Airport is underway. Construction of Mega Airport will cost 0 230 million. The project will be important for Gwadar Port and the city. C-Pack Authority Asim Saleem Bajwa

4.5.3.4 Frame: Benefits/Development/Roads Positively discusses road or railway projects in Pakistan re- lated to CPEC.

Example: ECNEC approves Main Line 1 Project ML-1 is biggest #CPEC project. In this project, Railway Track from Karachi to Peshawar will be doubled & upgraded. Financing Cost: $6.8 Billion Length of Railway Track: 1872KM Establishment of Dry Port in Havelian & Upgrading Walton Academy

4.5.4 Frame: Benefits/Investment

Positively emphasizes the size of investments coming with CPEC as part of projects or in general.

Example: CPEC will bring more than 82 Billion Dollars worth of investment from China alone in Pakistan. . .

4.5.5 Frame: Benefits/Jobs

Enphasizes the number of new jobs added as a result of CPEC. This also includes internships but not education or training initiatives.

Example: 1100 plus Jobs of various categories opened for a recently started #CPEC Project by Shanghai Electric at Thar Block-1.Locals to get preference subject to criteria/qualification. Plz apply on sites/addresses as on the flyer attached. #cpecmakingprogres

4.5.6 Frame: Benefits/Trade

Discusses trade in Pakistan as a result of of CPEC.

Example: Asim Bajwa is paving a wat for shining Balochistan & emerging Pakistan as a Chairman CPEC Authority The day is not far when Pakistan will be epicenter for world trade #AsimBajwaOurPride

References

1. [1] Dallas Card et al. “The Media Frames Corpus: Annotations of Frames Across Issues”. In: 2015. url: https: //www.aclweb.org/anthology/P15-2072.pdf.
2. [2] Saif M. Mohammad et al. “SemEval-2016 Task 6: Detecting Stance in Tweets”. In: 2016. url: https://www. aclweb.org/anthology/S16-1003.pdf.

Annex C: CP6 Belt and Road in East Africa Frame Annotation Guidelines

Task

Annotators will review texts from either Twitter, YouTube, or the Jamii forums and assign them with one of 3 stance values and 22 frames. They will also mark if the text is off-topic. Annotation will be done in batches, and we will periodically check for agreement and check in with annotators to discuss adding, removing, or modifying frame labels.

Off-topic label

- OT (use when the text is not related to both China and Africa)

Stance values:

- anti (anti-China)
- ? (Neutral/unknown)
- pro (pro-China)

Frame Values:

- Corruption
- Legal
- COVID
    - COVID/assistance
    - COVID/responsibility
- Debt
- Economy
- Infrastructure
- Jobs
- Land
- Investment
- Autonomy (previously "power")
- Quality
- Travel
- Prejudice
- Environmentalism
- Culture
- Politicians
- Cooperation
- Mistreatment
- Trade
- Surveillance
- UN

Marking Texts as Off-Topic

If a text is clearly not related to China and Africa, write "OT" in the "off-topic" column for the row. If the text is related to China and Africa, or if you can't tell because of missing context, don't write anything in that column. If a text is off-topic and not related to China and Africa, it should still receive a stance label, usually the neutral/unknown stance "?". It is possible for an off-topic text to have a pro or anti stance if the text mentions China but not any African countries. However, some texts that look off-topic may actually be from Chinese state-affiliated accounts with a focus on Africa.

As an example, this tweet doesn't mention China but it's from the official account of the Chinese Embassy to Kenya:
"#HelloAfrica XVI: Feature: Botswana farmers make fortune through quail breeding http://xinhuanet.com/english/2021-03/14/c_139809498.html"

The annotation set does include tweets from the reply chain, so it is expected that some examples will not be entirely relevant to the scenario.

Stance Value Descriptions and Examples

During annotation, please fill the "stance" column of each row with one of these values. Each row you review **must** be labeled with a stance, even off-topic texts.

- Anti (anti-China)

  Expresses negative attitude toward Chinese companies or government

- ? (neutral/unknown)

  The wording sounds neutral or objective, the stance is not clear, it's off-topic, or it is unknown

- Pro (pro-China)

  Expresses positive attitude toward Chinese companies or government

Frame Value Descriptions and Examples

Frames are labels that show what kinds of issues are mentioned in the text. A text that talks about court cases should get the "legal" frame, one that talks about debt should get the "debt" frame, etc. As you annotate, write the names of any frames that are related to the text in the "frame" column. A text might express one frame, multiple frames, or it might have none. If the text expresses no frames, leave the column empty. If it expresses multiple frames, write them in a comma-separated list (e.g. "corruption, legal").

Frame	Description	Example
Corruption	Discussion of bribery, embezzlement, or other forms of corruption	"@ChineseZimbabwe This is what you do to corrupt people in China, yet you want to apply a different standard in Zimbabwe because you are benefiting from the plunder of Zimbabwe's natural resources by the ZANUPF elites! It is corruption not sanctions that has killed Zim!"
Legal	Courts, legal issues, laws, lawyers, or legislation	"Kenya's new China-funded railroad is already uneconomical and a huge debt burden. Now a Kenyan court rules that the underlying contract was secured illegally, without the required public tender. No wonder it's another Belt and Road Initiative debt trap."
COVID	Coronavirus and related topics such as vaccines	"Chinese Ambassador to Kenya Wu Peng has revealed that Kenyans were mistreated in China as a result of them not maintaining social distance"
COVID/assistance (Chinese COVID assistance)	Chinese COVID assistance such as donations of protective equipment or doctors	"Kenya receives 100,000 face masks, 20,000 Covid-19 test kits and medical use protective suits from Chinese tycoon Jack Ma."
COVID/responsibility (China responsible for COVID)	Asserts that China is responsible for the COVID pandemic.	"In Kenya, we must also now sue China for damages and wrongful cause of deaths because of #COVID2019 Germany sends China £130billion bill for 'coronavirus damages' – sparks fury in Beijing"
Debt	Debt or loans, where the lender is China and the lendee is an African nation.	"After Tanzania, Kenya requests China for debt relief. Its critical to note that China holds 62% of #AfricanDebt. #Ethiopia PM said, "Africa is forced to choose whether to repay debt or redirect resources to save lives". China has used pandemic to push Africa deep into #DebtTrap"
Economy	Economic issues such as inflation and GDP	"The cabinet warning Cabinet warns people spreading fake news on COVID-19. Agencies ordered to pursue & arrest rumour mongers. Cabinet briefed on impact of COVID-19 on economy.Kenya to host Africa CDC after request by AU and WHO. Chinese gov't to build Ksh.8.3B 21 storey office"
Infrastructure	Roads, railways, ports, or other infrastructure.	"Kenya's new China-funded railroad is already uneconomical and a huge debt burden. Now a Kenyan court rules that the underlying contract was secured illegally, without the required public tender. No wonder it's another Belt and Road Initiative debt trap."
Investment	Chinese investment in Africa	"Liao Shen Industrial Park, invested by Chinese, has transformed Kapeeka into a manufacturing hub producing finished products that mostly go to export markets. As more are expected, people should pay a visit to Kapeeka indeed!"
Jobs	Jobs and employment.	"With investors avoiding Kenya I laugh every time someone talks about creating millions of jobs like this not a clay job that you wake up & mould. The kickback money is buying high end cars &

		homes,harambees, unless you'll be eating Chinese the promised jobs mtaonea viusasa.“
Land (Land Ownership)	Issues related to land ownership.	“Kenyans, for every 20 feet cargo container ferried by the SGR, Kenya looses $170. What's more, we pay the Chinese Sh 50 million daily for a project that has killed Kenyan businesses, jobs, livelihoods & benefits only a few land owners at the Naivasha & Nairobi dry ports! “
Autonomy	Topics related to autonomy, colonialism, imperialism, or exploitation.	“The new colonialists:Beijing mining boss 'shot Zimbabwe miners in pay row' “
Quality	Discussion of quality of goods from China, or counterfeit products from China.	“'No guarantee for China product' Tanzania proves it in a Royal way. They randomly test some animals for ChinaVirus and found positive!”
Travel	Travel between countries, usually by flights. Often in the context of COVID spread.	“@kipmurkomen @SenMutula @moseskajwang this is the kind of urgent stuff you should be debating in Senate. How did a plane from China get clearance to land in Kenya? What is self quarantine? Too many questions? The Health CS should resign.“
Prejudice	Discussion of racism, xenophobia, or other forms of prejudice such as religious discrimination.	”I am not happy at all with the treatment of Africans in China. This is pure RACISM and has to be called as such. We have to protect our people. @DrAlfredMutua @africansinchina @OurMayors @cgtnafrica @CSISAfrica @USAIDAfrica @AsstSecStateAF @LeMonde_Afrique @AfricaACSS “
Environmentalism	Discussion of environmental, pollution, national park preservation, or wildlife issues.	“Uganda: 37 Chinese nationals arrested for unlawful possession of illegal wildlife incl tortoises and pangolin scales and close to 2000 sim cards”
Culture (Cultural Connections and Soft Power)	Efforts for cultural connection or soft power between China and African nations. Connections between the countries that are not economic, political, or military.	“Imani Abdulrahman Haji of #Tanzania, a teacher working with the Confucius Institute at Dar es Salaam University, reviewed her journey of learning Chinese and the choice of becoming a disseminator of Chinese culture.”
Politicians	Discussion of people holding political office in African countries such as presidents, senators, etc.	“MT KENYA GETS NEW KINGPIN. Andrew Ngirici Forms The Citizens Convention Party (CCP) which will take Mt. Kenya by Storm and the whole country. SAUTI YA RAIA. “
Cooperation	Cooperation between China and African countries.	“"Good friends, good brothers and good partners." China, Zimbabwe mark 40 years of diplomatic relations, win-win cooperation. The 40 years have seen the two countries standing together not only on the political front, but also in many other fields “
Mistreatment	Description of mistreatment or harm of individuals or particular groups of people such as Kenyans, Africans in general, Chinese people, or religious groups.	”Chinese Ambassador to Kenya Wu Peng has revealed that Kenyans were mistreated in China as a result of them not maintaining social distance”

Trade	Trade between China and African countries.Topics include import/export and cargo shipments.	"According to FACT-CHECK Tilapia and frozen mackerel makes up 85% of Kenya's 2016 fish imports, with 68% coming from China. In 2017 China remained Kenya's top source of fish imports. Other fish product imports included tuna, Nile perch, sardines, fish feed, and fish waste."
Surveillance	Discussion of spying or surveillance by China.	"You tell Zimbabweans that Mnangagwa has paid the Chinese millions of USDs to build a snooping cybersecurity centre headed by an airforce Group Captain, to spy on them; they respond with lectures on end-to-end encryption & related crap. No wonder why Zimbabwe is in the doldrums!"
UN	Votes, meetings, or other issues related to the United Nations.	Example: "No joke: China now chairing a U.N. meeting of dictatorships to condemn sanctions against them (for their human rights abuses) as contrary to international law, with reps from Cuba, Syria, Belarus & Zimbabwe."