

# **AI-Enabled Social Cyber Maneuver Detection and Creation**

**Matthew Hicks**

CMU-S3D-25-103

April 2025

Software and Societal Systems Department  
School of Computer Science  
Carnegie Mellon University  
Pittsburgh, PA 15213

**Thesis Committee:**

Kathleen M. Carley, Chair

Patrick S. Park

Mohamed Farag

David M. Beskow (United States Military Academy)

*Submitted in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy in Societal Computing.*

Copyright © **Matthew Hicks**

The research for this paper was supported in part by the US Army project Scalable Technologies for Social Cybersecurity (W911NF20D0002) and by the Office of Naval Research under project Minerva-Multi-Level Models of Covert Online Information Campaigns (N000142112765) and Scalable Tools for Social Media Assessment (N000142112229), and by the center for Informed Democracy and Social-cybersecurity (IDeAS) and the center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research, Army, or the US Government.

The appearance of U.S. Department of Defense (DoD) documents and visual information does not imply or constitute DoD endorsement.

Keywords: social-cyber security, social network analysis, BEND framework, large language models, synthetic social media



## Abstract

As social media platforms have become central to information dissemination, influence operations, and narrative shaping, understanding their role within the broader information environment is increasingly vital. The BEND framework offers a structure for analyzing online influence by identifying social-cyber maneuvers.

The BEND framework was previously operationalized at the message and individual levels. In this thesis, I operationalize the BEND framework at the population and effects levels, integrate both sets of work, and align them with U.S. military doctrine and training. In doing so, I identify the critical need for complex, realistic, and scalable social media training environments.

To meet this need, I introduce the AI-Enabled Scenario Orchestration and Planning (AESOP) tool, which enables planners to create training scenarios that specify events, actors, social media platform accounts, and narratives. AESOP generates synthetic templates associated with the scenario and accompanying news articles, media content, and URLs.

I then present SynTel and SynX, agent-based simulation and generation tools. These tools consume AESOP-generated synthetic templates and, with support from external large language models, produce realistic and interactive synthetic social media data for X/Twitter and Telegram. These simulations replicate influence ecosystems at scale.

Finally, I propose and validate a novel effects-based approach to detecting BEND maneuvers within topic-oriented groups. This technique is applied to real-world datasets to link maneuver effects to broader campaign impacts.

Together, these contributions enhance our capacity to detect, evaluate, and train against influence operations — making BEND a practical analysis framework for the information environment.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Overarching Thesis Goals . . . . .	1
1.2	Overview of Chapters . . . . .	2
<b>2</b>	<b>Data and Tools</b>	<b>3</b>
2.1	Data . . . . .	3
2.2	Tools . . . . .	4
2.2.1	Existing tools . . . . .	4
2.2.2	Tools developed for this thesis . . . . .	4
<b>3</b>	<b>Background and Related Work</b>	<b>6</b>
3.1	Social Cyber Security . . . . .	6
<b>4</b>	<b>US Military Doctrine</b>	<b>9</b>
4.1	Research Questions . . . . .	9
4.2	Doctrinal Synthesis . . . . .	10
4.3	A Training Requirement . . . . .	12
<b>5</b>	<b>AI-Enabled Scenario Orchestration and Planning (AESOP)</b>	<b>14</b>
5.1	Research Questions . . . . .	14
5.2	A Training Approach . . . . .	15
5.3	AI-Enabled Scenario Orchestration and Planning (AESOP) Tool . . . . .	18
5.4	AESOP Outputs . . . . .	34
5.5	Future Work . . . . .	36
5.6	Conclusions . . . . .	37
<b>6</b>	<b>Synthetic Social Media Creation</b>	<b>39</b>
6.1	Research Questions . . . . .	39
6.2	Synthetic Generation Approaches . . . . .	40
6.3	SynX . . . . .	42
6.4	SynTel . . . . .	52
6.5	Validation . . . . .	54
6.5.1	Experiment . . . . .	57
6.5.2	Results . . . . .	58

6.6	Future Work and Limitations . . . . .	63
6.7	Conclusions . . . . .	64
<b>7</b>	<b>BEND: Effects-based detection</b>	<b>65</b>
7.1	Research Questions . . . . .	65
7.2	BEND-Effects Definitions . . . . .	66
7.3	Temporal Community Detection via Network Intersection and Clustering . . . . .	77
7.4	Results . . . . .	80
7.5	Future Work . . . . .	80
7.6	Conclusions . . . . .	81
<b>8</b>	<b>Conclusion</b>	<b>84</b>
8.1	Theoretical Contributions . . . . .	84
8.2	Methodological Contributions . . . . .	84
8.2.1	Effects-Based Detection of BEND Maneuvers . . . . .	85
8.2.2	SynTel and SynX: Agent-Based Social Media Generators . . . . .	85
8.2.3	AESOP: AI-Enabled Scenario Orchestration and Planning . . . . .	85
8.3	Application Contributions . . . . .	85
8.4	Limitations . . . . .	86
<b>Bibliography</b>		<b>87</b>
<b>A</b>	<b>AESOP Products</b>	<b>94</b>
<b>B</b>	<b>Synthetic Social Media Creation for Social-Cybersecurity Training</b>	<b>109</b>
<b>C</b>	<b>JSON Schema for Data Standard</b>	<b>138</b>
<b>D</b>	<b>Large Language Models and Synthetic Social Media Generation</b>	<b>163</b>

# Chapter 1

## Introduction

### 1.1 Overarching Thesis Goals

Social media platforms have an immense impact on information dissemination, influence operations, and narrative shaping. This provides an avenue for interested actors to influence public opinion, manipulate people, and even conduct war by other means.[10] Despite the growing importance of these dynamics, existing analytical frameworks and doctrinal tools — particularly within the U.S. military — have struggled to adapt. While other domains of military planning apply rigorous models and systematic assessments, social media often remains under-analyzed, poorly integrated, or misunderstood in operational planning and in training.

This thesis seeks to bridge the disconnect between emerging influence analysis frameworks and traditional U.S. military operations by synthesizing current doctrine with BEND - a social-cybersecurity framework - and providing practical tools for integrating social media into existing training exercises:

- AESOP (AI-Enabled Scenario Orchestration and Planning), a standalone Python-based application, empowers planners to construct, edit, and generate social-cyber scenarios grounded in real-world dynamics using configurable LLMs and structured inputs.
- SynTel and SynX, agent-based generators for Telegram and Twitter/X respectively, combine traditional simulation logic with LLM-driven message creation to produce realistic, platform-specific datasets.

These tools lower the barrier to high-quality scenario generation and provide a replicable method for developing training-ready content tailored to a variety of domains, including military, emergency response, public health, and law enforcement.

Additionally, this thesis seeks to enhance our ability to detect influence maneuvers — specifically those captured by the BEND framework. While BEND maneuvers are, by definition, effects-based, prior detection approaches have focused primarily on inferring the intent of the actor behind the message. This thesis reconceptualizes the detection problem by shifting the analytical emphasis to the observable effects of a maneuver within networks and narratives.

By developing a new effects-based detection methodology and integrating it with prior cue-based methods (e.g., CUE+), this research hopes to offer a more comprehensive, empirically grounded means of identifying influence activity in real-world social media datasets.

## 1.2 Overview of Chapters

This thesis is remarkably linear, with a single "Golden Thread" running through it - creating an environment for training BEND. Chapter 3 introduces the major concepts of social cyber security - including the BEND framework and the current status quo. These provide the basis for envisioning what such an environment might - must - look like. Chapter 4 builds on this by providing an emerging practical application of BEND within the US Department of Defense (DoD). The conclusions from this chapter, namely that BEND training is required and demands realistic and complex social media datasets, provide the demand signal for a BEND training environment. In Chapter 5, I introduce the AI-Enabled Scenario Orchestration and Planning (AESOP) tool for the creation of social media exercise scenarios - the first major component of the BEND training environment. In Chapter 6, I introduce the second major component - LLM-connected agent-based synthetic data generators SynX and SynTel which take AESOP scenarios and output corresponding datasets. Finally, in Chapter 7, I propose and demonstrate an effects-based method for detecting BEND maneuvers that goes beyond the current intent-inference approach that relies on language cues. This enables the training audience to interact effectively with the training environment - identifying not just intended BEND maneuvers but also providing the ability to evaluate their effectiveness.

# Chapter 2

## Data and Tools

### 2.1 Data

Previous research has examined several large-scale datasets for BEND maneuvers using existing techniques.[14] [13] I will apply BEND maneuver detection again, both through the current methodology[13] and the extended effects-based BEND maneuver detection discussed later in this thesis.

Table 2.1: Corpus Summary Statistics

Corpus Topic	Time Period	# Messages	# Agents
Balikatan 2022	~2022-04-07 thru 2022-04-14	2,372	1,308
Nice, France Terrorist Attack 2020	~2020-10-16 thru 2020-11-09	612,257	221,200
Synthetic Scenario	~2034-05-01 thru 2034-05-10	19,719	91

**Balikatan 2022** Balikatan is an annual bilateral military exercise between the United States and the Philippines. This data was collected from April 7 to 14 April, 2022, based on the keyword "Balikatan". Approximately 2,300 messages

**French Attack in Nice 2020** An Islamic extremist attacked and fatally wounded three individuals inside a Roman Catholic church in Nice, France, on October 29, 2020. A research team from Singapore collected two weeks worth of Twitter data from the Fall of 2020. I focus on the data from October 26 to November 4, 2020, covering before and after the event. Approximately 380,000 messages.

**Synthetic dataset** This dataset consists of X/Twitter API v1 data synthetically created with SynX (see Tools) and covering a Balikatan-like scenario that occurs from 01 May to 10 May 2034, which was constructed with AESOP (see Tools). Approximately 19,700 messages.

## 2.2 Tools

### 2.2.1 Existing tools

#### ORA-Pro

Formerly, the Organization Risk Analyzer (ORA) - Professional version, now known simply as ORA-Pro. ORA-Pro is a dynamic network analysis and visualization tool. ORA-Pro can import Twitter, Telegram, and Reddit data for detailed analysis.[22]. This thesis takes advantage of the built-in stance detection, Leiden grouping, and social metrics as well as the BEND maneuver detection through CUES.

#### NetMapper

NetMapper processes text to identify concepts and their network relationships. It uses dictionaries and custom parameters to enrich text before extracting concepts, which it then links together to create either semantic or conventional meta-networks.[22] This thesis relies on NetMapper for extracting CUES from social media corpora before importing the CUES into ORA-Pro.

#### OpenAI Models and API

This thesis required large amounts of LLM interaction and the OpenAI API was user friendly and easy to integrate with Python.[62] I used a GPT4o mini model (gpt-4o-mini) for text and DALL-E model (dall-e-2) for images.

#### Local Large Language Model

Unfortunately, OpenAI models refused to respond properly to some requests for negative BEND maneuvers or to some negative topics due to guardrails. In those cases, I ran a large language model locally. For this thesis, I used mixtral-8x7b based on its effective responses, lack of guardrails, and small size.[6]

### 2.2.2 Tools developed for this thesis

#### AI-Enabled Scenario and Orchestration Planning (AESOP) Tool

AESOP allows Information Environment planners to develop social-cyber exercise scenarios from scratch or develop social-cyber vignettes for integration with existing scenarios. It is a standalone GUI coded in Python with PySide6 that leverages external LLMs.[67] AESOP was coded specifically to support the research in this thesis. A detailed description of AESOP can be found later in this thesis.

#### SynTel and SynX

SynTel and SynX are agent-based simulators, programmed in Python, that interact with LLMs to produce Telegram and X data, respectively. They were coded specifically to support the research

in this thesis. A comprehensive breakdown of how SynTel/X are constructed and operate can be found later in this thesis.

# **Chapter 3**

## **Background and Related Work**

### **3.1 Social Cyber Security**

Looking at cyberspace through the lens of warfare is not new. Interactions between adversaries within cyberspace have often been referred to in military terms of attack and defense [26]. Cyberspace simulations have been used to model these conflicts, often closely emulating current physical military doctrine [35]. However, these simulations focus primarily on the cyber-terrain itself - accurately deducing that terrain has a large impact on the outcome of conflict [35]. However, just as the physical domain of warfare stretches into the digital space, so too does the information environment. This is social cyber security [21] [20]. Social cyber security welds the methodologies of the social sciences with the need to identify, assess, and counter the impact of information maneuvers.[20]

This field is often claimed/mishandled by multiple interested parties. In the US Department of Defense (DoD), there is not only Joint (all-service) doctrine[45] addressing it but there is also Service doctrine[29] and within services there is Branch doctrine[31] - all sticking a finger into the mess that is the social cyber security component of the much broader Information Operations.

Although recently identified as an academic field and still nascent within the Department of Defense, the need for a framework to scaffold understanding of these issues is not new and has led to the rise of various contenders. These include Ben Nimmo's 4 D's - dismiss, distort, distract, and dismay - focusing on Russian propaganda techniques.[58] Also, the ABC framework developed by Camille Francois[37], which looks at the Actors, Behaviors, and Content of a disinformation campaign and its successors ABCD[8] and ABCDE.[63] Finally, the SCOTCH framework - focusing on Sources, Channels, Objectives, Targets, Composition, and Hooks - was brought forward by Blazek in 2021.[16]

Amid this crowded field lies the BEND framework. BEND provides a framework for discussing social-cyber interactions using narrative and network structures, but borrows the idea of informational maneuver from maneuver warfare [12]. BEND is shorthand for the social-cyber maneuvers: back, build, bridge, boost, engage, explain, excite, enhance, negate, neutralize, narrow, neglect, dismiss, distort, dismay, and distract. These maneuvers and their definitions are taken from Beskow and Carley's 2019 work Social cybersecurity: an emerging national security requirement [12] as refined and validated by Blane et al. in 2022 [14] and later in Blane's thesis

work.[13] BEND arguably stands out from other frameworks for several reasons.

First, BEND is detailed enough to provide leaders with a lexicon capable of expressing their specific desires. Second, it focuses on communicating a general understanding of the intent and effects of the information maneuver without becoming a low-level enumeration of the tactics, techniques, and procedures used in the execution of those maneuvers. Both of these distinctions are more than just attempts at carving out a niche for BEND - there exists a broad requirement for discussions of the intent and effects of maneuver without getting bogged down in the execution - and it persists across domains. Here, BEND lives in stark contrast with SCOTCH, a more detailed framework that provides a much richer execution scope for maneuver at the cost of brevity. SCOTCH requires a full operations order, while BEND is focused more on evoking just the broader tactical task. ABCDE is similar in its lack of brevity - requiring five paragraphs for full enumeration - but does manage to keep all of it at the non-execution (operational) level. Third, BEND is not limited to disinformation or to just the narrative side of information maneuver - in contrast with Nimmo's 4Ds. BEND fully categorizes maneuvers through narrative and network space and positive and negative impact.[13]

Lastly, the maneuver portions of BEND can - and should - be used to enrich any social-cybersecurity framework. It is already present in Nimmo's 4Ds but is necessary as an extension to cover positive and network maneuvers. It fits well within the Effects (E) of the ABCDE framework - giving concise scaffolding to an otherwise bulky framework - and it provides a shorthand for grouping identified SCOTCH enumerated campaign events - without needing to specify the execution pathway of each. Because BEND is not just focused on being an identifying framework, but also on short-handing important motifs within social-cybersecurity it offers utility everywhere.

BEND is not just the expression of maneuvers, nor is it just the categorical formatting for them, it is also a methodology for extracting maneuver intent and effects from the information environment. Currently, the CUE+ method as outlined by Blane [13] is the cutting edge in BEND maneuver detection. This method has seen several iterative improvements - first in Uyheng et al. in 2020[68], then by Blane et al. in 2022[14], then Alieva, et al. in 2022[9], before Blane adopted the the current method in her thesis work.

In this method, linguistic cues are extracted from message text using NetMapper software.[22] These NetMapper cues are a proprietary blend of concepts that represent a message's sentiment and the author's emotional state.[20] These particular cues - now referred to as CUES - are then loaded into ORA-Pro[22] - a network visualization and analysis tool - where they are mapped to the original message. ORA-Pro is able to use these CUES, along with supplemental network information about the message sender, to provide a report that identifies BEND maneuvers associated with messages and actors - i.e. who is trying to do what to whom.

This is necessary to identify BEND maneuvers within messages. However, the BEND maneuvers themselves, even in the extended definitions provided by Blane, are not intent-based - who is trying to do something - but impact-based - what actually happens. Indeed, the BEND maneuver descriptions and illustrative impacts provided by Blane are diagrams or illustrations of the effects the BEND maneuver will have. For instance, the diagram for the Boost maneuver is shown in Fig. 3.1

There is no mention of the content required within the message that caused the boost - the implication is that the message is defined by the impact it had. Blane goes on to derive message

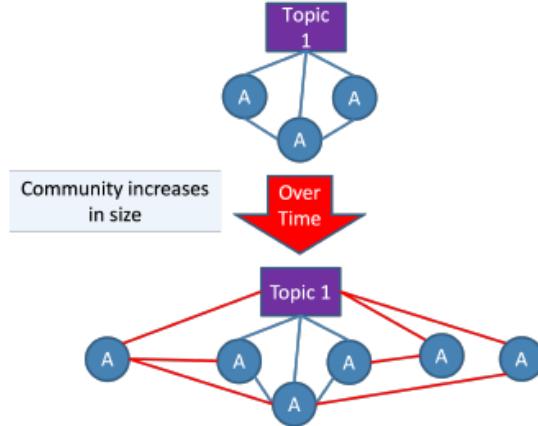


Figure 3.1: Excerpt from Blane's thesis work "Social-Cyber Maneuvers for Analyzing Online Influence Operations".[13]

content requirements based upon CUES. Again, this is necessary to identify which messages are attempting to conduct which maneuvers. However, assessing the effects of the messages - detecting BEND not within the messages but in the effects these messages had on their targets - is missing. This represents a fundamental disconnect between the state of the art in BEND detection (CUE+ based) and the actual intended use of BEND. The crux of this thesis aims at bridging this disconnect.

# **Chapter 4**

## **US Military Doctrine**

### **4.1 Research Questions**

As the Department of Defense continues to evolve its approach to operations within the Information Environment (IE), it faces significant challenges in conceptualizing, organizing, and executing influence efforts—especially on social media. The BEND framework offers a structured, effects-based lens through which influence and manipulation can be analyzed and operationalized. However, for BEND to serve as a useful tool in defense planning and operations, it must align with existing doctrine and be expressed in language and formats familiar to military decision-makers.

This chapter examines how the BEND framework can be nested within current U.S. military doctrine and proposes doctrinally consistent products that apply BEND to support decision-making and planning processes. The aim is not only to demonstrate BEND’s conceptual fit, but to create doctrinally relevant outputs—such as overlays, situation templates, running estimates, and targeting inputs—that enhance the military’s ability to visualize and respond to the social media dimension of the IE.

**The key research questions for this chapter are:**

- How does the BEND framework fit into current military doctrine?
- How can BEND enhance current information environment analysis?

To answer these questions, the chapter synthesizes numerous doctrinal sources, including Joint Publications, service-specific field manuals, and defense instructions, while highlighting existing gaps in how social media is conceptualized and managed. It also draws from practical applications in training environments, particularly the Project OMEN series, to show how BEND can inform both planning and execution through standardized outputs.

Ultimately, this chapter demonstrates that BEND can serve as a doctrinal bridge—linking narrative analysis, social media maneuver detection, and operational planning in a coherent, scalable way. It concludes by identifying a key area of opportunity within the DOTMLPF framework: the urgent need for realistic, doctrinally informed training to prepare military personnel to operate effectively within the social-cyber domain.

## 4.2 Doctrinal Synthesis

Current military doctrine on the social media aspect of the Information Environment (IE) is scattered between dozens of manuals and instructions and is encumbered by issues of both authority and ability. Current doctrinal examples that address the IE include:

- JP 3-13 Information Operations[45]
- NWP 3-13 Navy Information Operations[33]
- ADP 3-13 Information[32]
- AFDP 3-13 Information in Air Operation[29]
- CJCSI 3210.01C Joint Information Operations Proponent[23]
- DODI 3600.01 Information Operations[28]
- ADP 5-0 The Operations Process[30]
- MCWP 3-32 Marine Air-Ground Task Force Information Operations[40]
- JP 2-01.3 Joint Intelligence Preparation of the Operational Environment[44]
- ATP 2-01.3 Intelligence Preparation of the Battlefield[31]
- JP 3-60 Joint Targeting[43]
- JP 3-61 Public Affairs[46]

This is neither a comprehensive list of all applicable doctrine nor does it include those manuals which retain either a Secret classification or Controlled but Unclassified Information (CUI) identifier. Throughout these doctrinal examples the IE is defined as "The aggregate of individuals, organizations, and systems that collect, process, disseminate, or act on information." [45] Unfortunately, because of the importance of the IE it is often discussed as the "Information Domain" - something not explicitly found in US doctrine. In none of these manuals is "domain" officially defined[4]; however, JP 1 Doctrine of the Armed Forces of the United States does discuss the "the physical domains (air, land, maritime, and space); the information environment (which includes cyberspace); and political, military, economic, social, information, and infrastructure (PMESII) systems and subsystems." [47] The fact that information is both an environment akin to the physical domains and a separately listed system should speak to its importance. The implication is that there exist separate regions marked by distinct physical characteristics (land, air, sea, etc.) and a mostly intangible information region. This is borne out in discussions of domains - they often include the information domain despite its lack of doctrinal pedigree[7].

The doctrine elsewhere includes the information domain as an instrument of national power[47]. These instruments of national power are laid out as a part of the DIME framework which defines diplomatic, informational, military, and economic instruments - more recently expanded to include finance, intelligence, and law enforcement (DIME-FIL)[65]. Indeed, the integration of the information "domain" or environment was the driving force behind the Department of Defense's Joint All Domain Command and Control (JADC2) initiative[5]. Whether a domain or an environment, information holds relevance equal to any of the physical domains.

As social media becomes increasingly important within the Information Environment[7], the BEND framework provides a solution for proper analysis and lexicon across warfighting func-

tions. There are a wide variety of actions and actors within the DoD concerned with information operations. These include Public Affairs Operations (PAO) - "provide accurate and timely information" to the public[46] - Military Information Support Operations (MISO) - "influence the attitudes, opinions, and behavior of foreign target audiences"[46] - and Military Deception (MILDEC) - "deliberately mislead adversary... decision makers"[46]. All of these benefit from the common lexicon of tactics and maneuvers provided by BEND.

Unfortunately, US Department of Defense (DoD) actions in the IE are also hampered by issues of authority and capability. These issues are highlighted most clearly when considering social media. The authority for the DoD to conduct "operations" within social media is limited by a wide array of both law and policy. Limiting factors include the Posse Comitatus Act[3], the Fourth Amendment[1], the Privacy Act of 1974[2], and numerous DoD directives and regulations. These factors severely limit the collection of information on US citizens and the conduct of narrative campaigns (beyond informational) that target US citizens. This thesis does not address solutions for these issues and assumes the DoD has already established or will establish the proper authorities to conduct analysis and response within the social media subsystem of the information environment.

While these authorities remain a legal question, the capability issue can be addressed. Decision makers need mappings of the information environment to know how to analyze it and respond within it. In other domains, leaders rely upon the Modified Combined Obstacle Overlay (MCOO)[44]. The land domain MCOO is perhaps the easiest to conceptualize. Obstacles, avenues of approach, key terrain, observation and fields of fire, and cover and concealment are graphically depicted atop a topographic map of the battlefield. This provides decision makers with a clear understanding of how the pieces of the land battle interact. The MCOO is further enriched by adding the enemy situation template (ENY SITEMP). The enemy situation template shows the disposition of known enemy positions overlaid on the MCOO, as well as the enemy's most dangerous or most likely course of action based on the enemy's doctrine. The lifting of this concept and application of it to the IE is not a new idea and others have posited the creation of a Combined Information Overlay [27]. This is encouraged by the description of the Consolidated Systems Overlay in JP 2-01.3 Joint Intelligence Preparation of the Operational Environment [44], see Fig. 4.1, which is unfortunately never explicitly applied to the information environment.

**The Social Media CSO/MCOO** BEND is key to combining the methodological rigor of the Modified Combined Obstacle Overlay (MCOO) with the systems perspective of the Consolidated Systems Overlay (CSO) in order to graphically depict the current state of the social media component of the information environment for decisions makers. The information environment needs a MCOO for the social media subsystem to feed the social media enemy situation template, the social media running estimate, the friendly social media situation, aid in course of action development, and provide meaning to an IE impact assessment.

Project OMEN has been a major force in developing examples of the application of BEND within the military. Project OMEN is a training scenario designed to educate players on social media analytics.[50] In work done in support of Project OMEN, I constructed products that match each of these areas from the BEND reports and outputs available. Beyond this, I will develop a social media running estimate template for information operations staffers that includes these

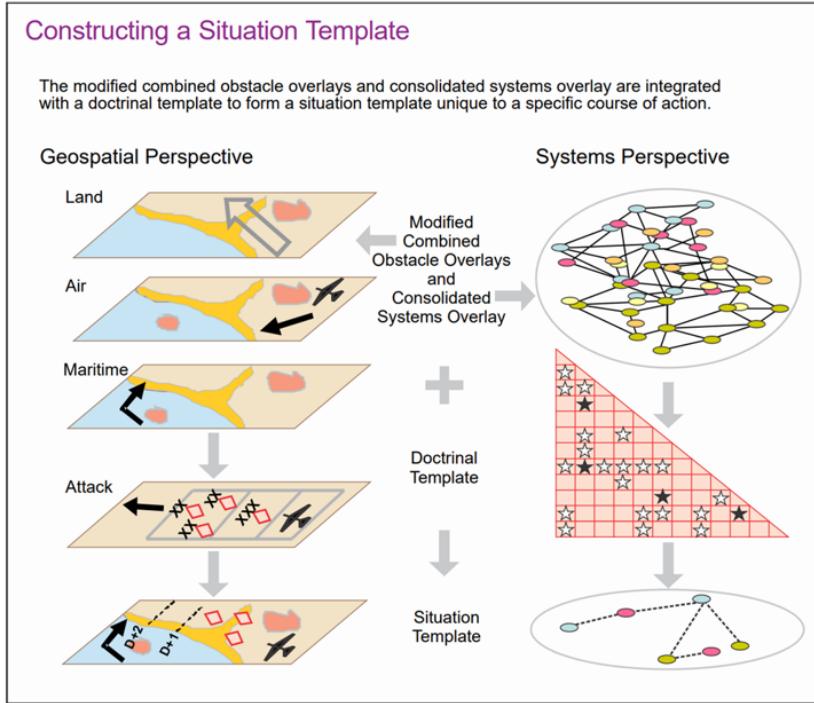


Figure V-2. Constructing a Situation Template

Figure 4.1: MCOO and CSO comparison from JP 2-01.3 Joint Intelligence Preparation of the Operational Environment

examples and accompanying narratives.

These products synthesize with current US military doctrine. They are aligned with the Joint Targeting Cycle (JTC)[43], Joint Intelligence Preparation of the Operational Environment (JIPOE)[44], and the Joint Operation Planning Process (JOPP)[48].

It is important to note that while social media is a critical component of the information environment, it is still only a subsystem. These products help scaffold the leaders and their staffs in understanding the role social media plays; however, social media is not a domain or an environment level consideration. The products outlined here reflect only a subset of those that would feed larger information environment planning processes and cycles. Samples of these products can be found in Appendix ??.

## 4.3 A Training Requirement

The DoD uses a framework called DOTMLPF to identify where changes need to be made or can be made in the development or fielding of new solutions, be they equipment, units, or even concepts. DOTMLPF is an acronym addressing capability across seven interrelated domains: Doctrine, Organization, Training, Materiel, Leadership and Education, Personnel, and Facilities. As a burgeoning concept within the DoD, social-cybersecurity is no different from any other capability. Thus far, Organization is assumed to be addressed as a nascent entity with the requisite

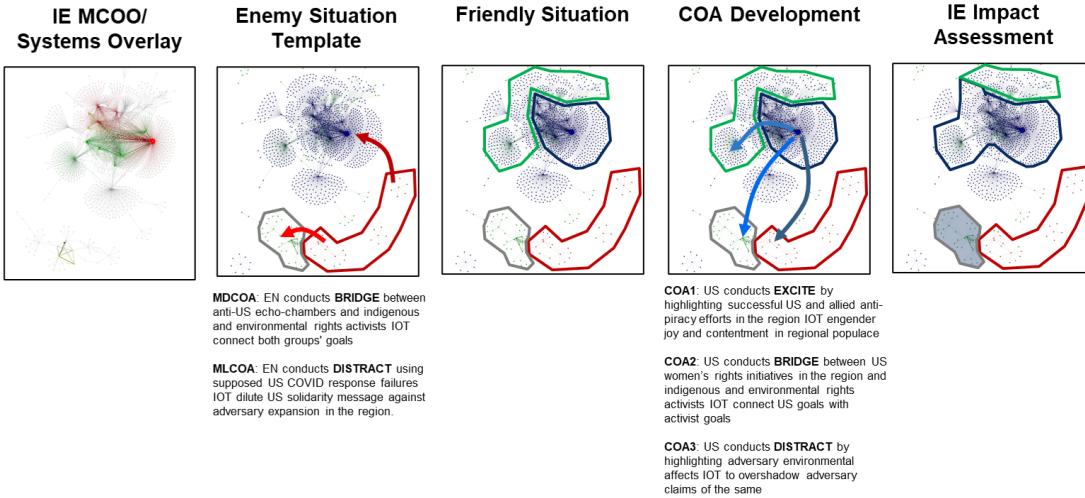


Figure 4.2: Information Environment products influence by BEND as a part of this thesis.

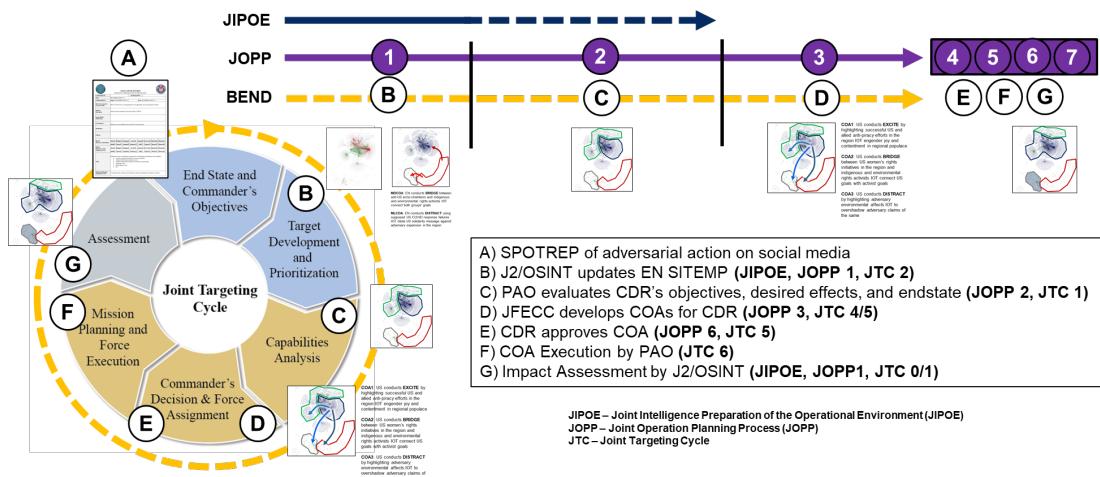


Figure 4.3: BEND integration with the Joint Targeting Cycle, JIPOE, JOPP.

authorities for conducting social media maneuver. Doctrine is addressed here, and while there is not yet clear understanding of particular doctrine, there is direction forward, which is sufficient for developing capabilities. Materiel, Leadership and Education, Personnel, and Facilities are well outside the scope of this thesis and require executive decision and legislative action. This leaves Training. There is an outstanding requirement for complex, realistic training within the social-cybersecurity landscape.

# **Chapter 5**

## **AI-Enabled Scenario Orchestration and Planning (AESOP)**

### **5.1 Research Questions**

Understanding the BEND framework, identifying BEND maneuvers and their effects, and even conceptually understanding the doctrinal application of BEND fall short of fully operationalizing BEND. What is required is realistic training on a realistic corpus where a training audience can apply these concepts. A point of clarification is required here because training - especially when artificial intelligence and large-language models are included in the discussion - can mean different things to different people. Training often refers to model training, in which a model learns a classification task from a training data set.[38] This is not the case here. Here, training involves the instruction and learning reinforcement of living, breathing humans on BEND and social media analysis techniques.

Effective training for social-cyber analysts requires realistic, relevant, and dynamic data environments. However, the practical constraints of using real social media data—including legal, ethical, and operational limitations—often render it insufficient or inappropriate for targeted training needs. Synthetic data provides a compelling alternative, but its utility depends entirely on how well it mirrors the complexity and nuance of real-world online behavior, both in content and in structure.

This chapter introduces the AI-Enabled Scenario Orchestration and Planning (AESOP) tool, a system developed as part of this thesis to generate training-ready synthetic data grounded in real-world dynamics. Before diving into AESOP’s design, this chapter begins by categorizing and evaluating the primary types of data available for social media training. Using criteria such as relevance, scalability, interactivity, and realism (both network and narrative), I compare real data, hybrid models, and various synthetic approaches to determine which best support different training objectives.

From this analysis, it becomes clear that synthetic data informed by real-world patterns—and built explicitly to support narrative and maneuver analysis—is uniquely suited for BEND training. AESOP operationalizes this insight by enabling scenario planners to design realistic and customizable information environments. Planners define actors, groups, events, narratives, and

supporting media, while AESOP scaffolds this creation process using large language models (LLMs) to reduce cognitive load and improve immersion. AESOP's outputs include both synthetic templates, which feed into generators like SynTel and SynX, and exercise-ready documents for participants and controllers.

By formalizing and standardizing the scenario design process, this chapter bridges the gap between high-level training objectives and low-level data generation, ensuring that synthetic social media corpora are not only technically robust but pedagogically effective.

**The key research questions for this chapter are:**

- How can we develop exercise training scenarios for the BEND framework?
- Can we extract a training scenario from real data without resorting to hand-crafting messages?
- How can we leverage AI/LLMs to enhance training scenarios and generate multi-modal BEND maneuvers?

## 5.2 A Training Approach

The idea of realistic training for social-cyber security is not new. Project OMEN has been steadily increasing the complexity of training scenarios for the DoD since 2021[50], culminating in February 2023 with a hybrid of hand-altered real-world Twitter/X data and synthetically generated Telegram data. This mix of data is a result of increasing demands for relevance, scalability, and realism in networks and narrative.

Table 5.1: Comparison of Data Types for Modeling Social Media Environments

	Relevance	Scalability	Interactivity	Realistic Networks	Realistic Narratives
Real Data	-	+	-	+	+
Hybrid: Hand alteration	o	-	-	o	o
Hybrid: Automated alteration	o	o	-	o	o
Current Synthetic	+	+	+	-	o
Synthetic Derived from Real Data	+	+	+	+	o

Skipping over the question of scenario design for now, it is conceptually possible to train individuals on BEND social media analysis on a wide array of social media corpus. However, data selection should always match training objectives. It is, therefore, helpful to categorize the options for data that can be used during training and evaluate them across five areas.

The first area is relevance. Relevance is meant to encompass two related concepts - adaptability and applicability. This may seem an odd pairing for a single category, but they are closely woven together. Data can be applicable - it may include the exact topics and events required for training - and therefore does not require reconfiguration, or it may be adaptable such that it can be made applicable for the training. Relevance is here a measure of both as it pertains to the content of the data, i.e. is the data already topical or can it easily be made topical?

The second area is scalability. It can be difficult to draw an analysis out of too little data, just as it can be challenging to handle too much data. Ideally, data will be of sufficient size to

meet particular training objectives - scalability is measure of how easy it is to shift a dataset to the correct quantity.

The third area is interactivity. Thus far, the corpus is discussed as a monolithic entity for the training audience. However, feedback is an important part of training - the training audience requires reinforcement from the training environment when they are making the "correct" action and negative feedback when they are making the "incorrect" actions. Interactivity is a measure of how easily a data set can be changed such that the training audience can see both the effects of their own actions within the corpus and evaluate these effects against desired outcomes.

The fourth and fifth areas are closely related and deal with the realism of the data in both the network and narrative domains. The data should closely reflect real-world social network constructs and real-world narratives, and both must vary appropriately per platform, per topic, etc.

With these criteria in hand, we can evaluate several approaches to data construction for training purposes. Major data approaches include the harvesting of real data, the hand alteration of harvested real data, the automated alteration of harvested real data, the bespoke creation of synthetic data, and using real-world data to scaffold the creation of synthetic data. In general, these represent gradient lines on a spectrum and are not meant to be comprehensive of all approaches.

In looking at these approaches against the criteria, we can evaluate their suitability for training. Real data is inflexible in its ability to adjust to changing training requirements or objectives. It is highly unlikely that the exact event required for training happened at exactly the right time in a real-world scenario that enables collection of a suitable dataset. It is already focused on an event and/or topic - which may or may not fit within the confines of the training requirement. It is definitionally non-reactive and will never reflect the actions taken by the training audience. However, it is scalable - just collect more - and the gold standard for realism from both a network and narrative perspective. Hand altered real data is somewhat reconfigurable to meet training demands - however, doing so prevents any level of scalability - it takes too long to alter large amounts of data - and the altering can reduce network and narrative realism. Automated alteration retains the benefits of hand altering but adds the ability to scale. Observed current synthetic environments are imminently reconfigurable and scalable but output unrealistic networks and weak narratives. These trade-offs are displayed in Table 5.1.

Toward this end, this thesis is concerned with using the analysis of real data to attempt to inform the construction of a tailored exercise scenario that includes defined actors, events, groups, and narratives to aid in the creation of synthetic data. The desired end result is a relevant, interactive, scalable, realistic dataset accompanied by corresponding products to enable training.

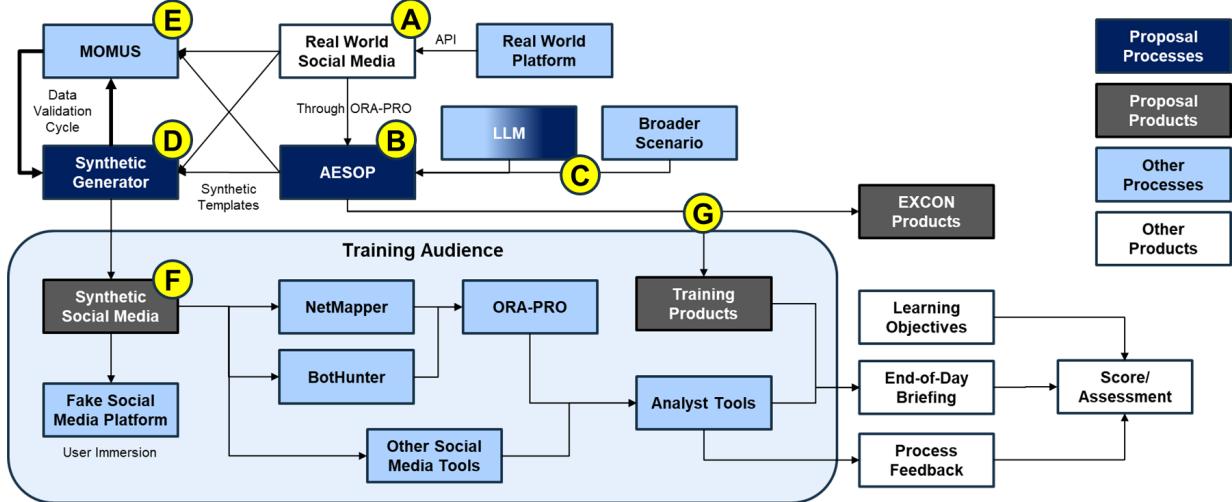


Figure 5.1: Project OMEN training flow.

In order to better understand how this might be accomplished it is helpful to look at the overall flow of an example training exercise. Fig. 5.1 shows the Project OMEN training flow - this training flow is generalizable to any social media analysis exercise.

- A) Real-world data is pulled from a social media platform using approved APIs and without violating the terms of service.
- B) Important network and narrative characteristics are drawn from real-world social media data using ORA-Pro. These characteristics are passed to a scenario planning tool called AESOP (AI-Enabled Scenario Orchestration and Planning).
- C) A scenario planner uses AESOP in conjunction with their understanding of the training objectives of the training audience to construct a scenario. AESOP leverages large language models to fill in the gaps between real-world characteristics and the desired scenario.
- D) AESOP outputs templates for actors, groups, events, narratives, and new stories that are passed to the synthetic generator. These templates represent both the "needles" and the "haystack" for the training scenario. The synthetic generator uses these templates to generate social media traffic - ostensibly creating the "needles" and hiding them in the "haystack."
- E) The synthetic social media is passed to an automated evaluation system for scoring. In this case a Netanomics system called MOMUS. There is a cycle of validation as the synthetic generator iterates against evaluation system - whether MOMUS or something else.
- F) The synthetic social media is given to the training audience for training. The training audience uses tools to analyze the data - hopefully finding the "needles" within the "haystack."
- G) AESOP also outputs two sets of documents along with the scenario templates. The two sets of documents are a list of events and hosted websites for exercise control personnel

and a set of baseline documents for the training audience that act as a breadcrumbs to orient them to the "haystack."

The scenario planning tool (AESOP) and the synthetic generator were constructed completely for this thesis. Outputs from work done as a part of this thesis include synthetic social media corpora, products for the training audience, and products for the exercise controllers.

## 5.3 AI-Enabled Scenario Orchestration and Planning (AESOP) Tool

AESOP allows Information Environment planners to develop social-cyber exercise scenarios from scratch or develop social-cyber vignettes for integration with existing scenarios. It is a standalone program coded in Python with a PySide6 GUI.[67] While AESOP was developed with US military exercise scenario development in mind, the principles and the tools can be applied to a wide swathe of exercises, including scenarios involving the health sector, emergency response, law enforcement, COOP and disaster recovery, etc.

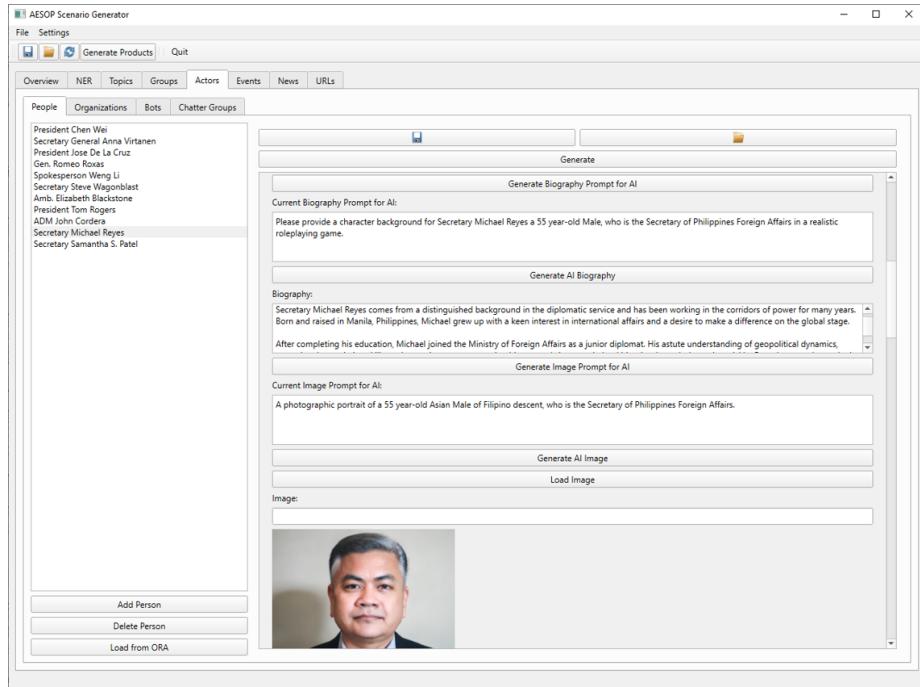


Figure 5.2: AESOP GUI Example. With an actor template being worked on.

AESOP leverages large language models (LLMs) to reduce planner load and increase realism and immersion for the training audience. Planners complete basic fields – such as date ranges and summaries – and AESOP develops an engineered prompt for a configurable LLM that is used to generate surrounding details. Planners can make additional changes to the prompt as required. Planners can also freely manipulate the details returned by the LLM. By default, AESOP reaches out to the OpenAI API[62] and uses a GPT4o mini model (gpt-4o-mini) for text and DALL-E

model (dall-e-2) for images. However, it is configurable to run against any LLM provider that has an OpenAI compatible API - such as the popular oogabooga/text-generation-webui.[61]

Planning a social media exercise is a complex topic that is beyond the scope of explanation for this thesis; however, understanding the basics is not difficult. The social media scenario will need non-training audience entities - sometimes called non-player characters (NPCs) - within the environment. The planners will want events to happen in the scenario in a time-sequence that is set or managed by the planners. The NPCs should hold opinions of their own on these events - and they should act based on these opinions. A social media scenario needs people (WHO) interacting about things (WHAT) in a certain way (HOW). The training audience will need to analyze the WHO/WHAT/HOW in order to intervene appropriately. There is obvious complexity missing from this brief explanation - the embedding of WHEN within both the WHO and the WHAT - the blending of the WHY with HOW, etc. But the WHO/WHAT/HOW explanation is sufficient to understand what is required for a scenario. To support scenario creation, therefore, AESOP is broken down into these three areas.

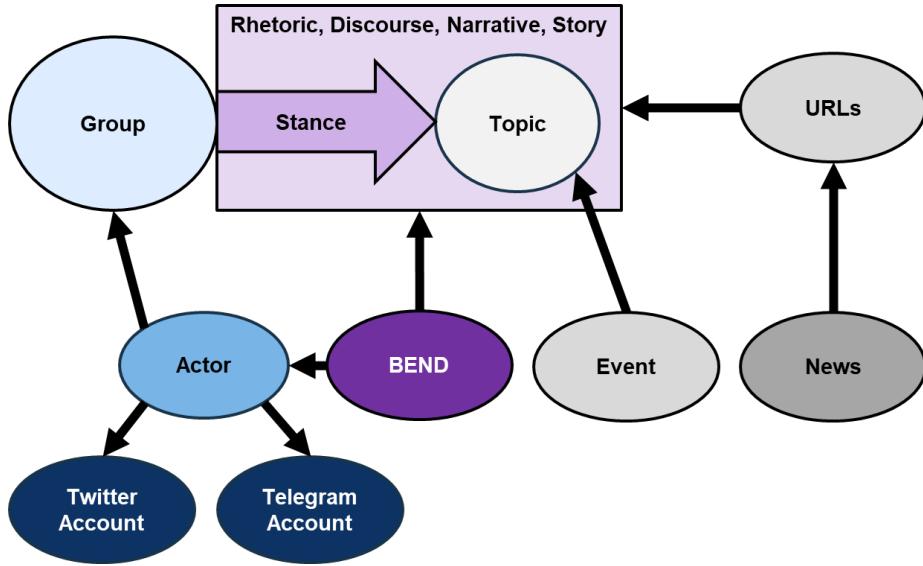


Figure 5.3: AESOP Relational Diagram. Blue is loosely defined as the WHO, purple is HOW, and gray is the WHAT.

## Actors

The WHO area includes groups, actors, and accounts. Actors are the independent actors within a scenario - this includes individual persons, organizations, and bots. These actors have names and biographies, cross-platform identity markers that distinguish them, and they have individual proclivities for certain BEND maneuvers.

### Person

- Name
- Leader Type [Political, Military, Organizational, Other]

- Title
- Organization
- Gender [M, F, Other]
- Race
- Nationality
- Age
- Entourage [Other Actor or Create New to establish relationships]
- Description
- Biography
- Image

### **Organization**

- Type [Government, News, Corporation, NGO, Political, International, Armed Forces, Charity, Non-Profit, Education, Interest-based, Other]
- Organizational Leader
- Description
- History
- Image

Each organizational type also populates a sub-list of attributes that are not enumerated here.

### **Bots**

Bots mimic either a person or an organization so they each include one set of the above attributes as well as:

- Type [Amplifier, News, Bridging, Repeater, Spam, Other]

The type then drives additional entries. For instance, a News bot has:

- Type of News [Regional, Industry, Political, Weather, Other]
- News Aggregator or News Producer
- Pink Slime
- Multiple Sources
- Retweeted accounts
- Common mentions
- Common places
- Keywords
- Hashtags

While this represents a large number of fields, AESOP scaffolds planners by filling in most of an actor's information based on just a few fields. Planners generally only need the first few fields (name, age, race, etc.) and a one-sentence characterization of the actor and then AESOP

leverages a large language model to complete the rest, including biographies and images. This is done through the use of intelligent prompting similar in concept to the children's fill-in-the-blank game Mad-Libs. These prompts are visible to the planner and configurable/editable if the default mad-lib prompt provides suboptimal results.

Finally, in order to support dynamic changes and interaction with the training audience, each actor has an Actor Topic Map (ATM). The ATM is a vector with a length equal to the number of topics (defined later in this chapter) and an opinion value from -5 to 5 approximating a Likert scale of approval for each topic.

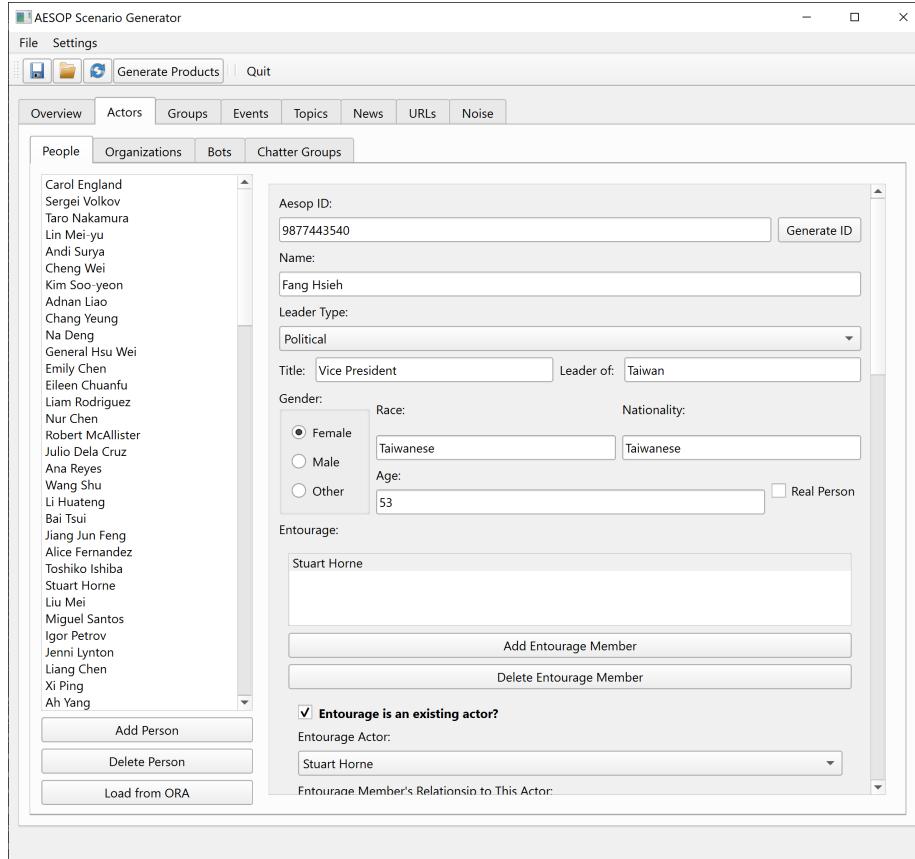


Figure 5.4: AESOP GUI for Actor creation.

## Accounts

Each actor is tied to one or more accounts. These accounts represent that actor's presence on a social media platform. AESOP supports the creation of both X/Twitter and Telegram accounts with shells for Facebook and Reddit accounts.

The full enumeration of what AESOP outputs for each account can be found in the data standards annex. As a representative example, this is a sample list of the fields for a X/Twitter account:

- Twitter ID

- Username
- Active dates
- Bot status
- Verified status
- Tweet distributions per day
- Account creation date
- Number of followers
- Number of following
- Number of original tweets per day
- Top topics
- Number of mentions per tweet
- Accounts to mention
- Number of retweets per day
- Number of quotes/replies per day
- Number of hashtags per post
- Top hashtags
- Top words
- Account daily active period
- Percent of tweets/retweets that are Positive/Negative/Neutral in sentiment
- Ratio of Text/Images/Video
- Additional identity markers unique to the platform

Again, this many decisions would be daunting for exercise planners - especially since, within the DoD, such planners are not expected to have any social media experience. AESOP scaffolds planners here as well. There are three options. Option one is default entries - AESOP prefills everything except the username with intelligent averages. The Twitter ID is auto-populated with a Snowflake compliant identification number, the active dates default to the active dates of the overall scenario, account creation is set to random dates before the beginning of the exercise, etc.[72] For a general account this option might be good enough.

Option two is to load an actor from ORA. The planner can use a real-world data set in ORA and export a Node of Interest Characterization report for a node that they want to replicate within the scenario. This exported file can be loaded into AESOP and will create an account with all fields populated to match the original real node - except the Snowflake ID, username, and active dates. This is most helpful for planners that want a certain character that has an archetype readily available in a real dataset.

The third option is to load a corpus of data directly into AESOP. For X/Twitter this would be an API v1 JSON file containing the desired tweets. AESOP will process these tweets with Latent Dirichlet Allocation (LDA), statistical analysis, and LLM summarization calls to fill out

all of the information as an average of the input data.[17] Again, except for the Snowflake ID, username, and active dates. In this way, planners are scaffolded towards defaults or choosing particular archetypes for their characters rather than creating bespoke accounts based upon platform-specific knowledge which they generally will not have.

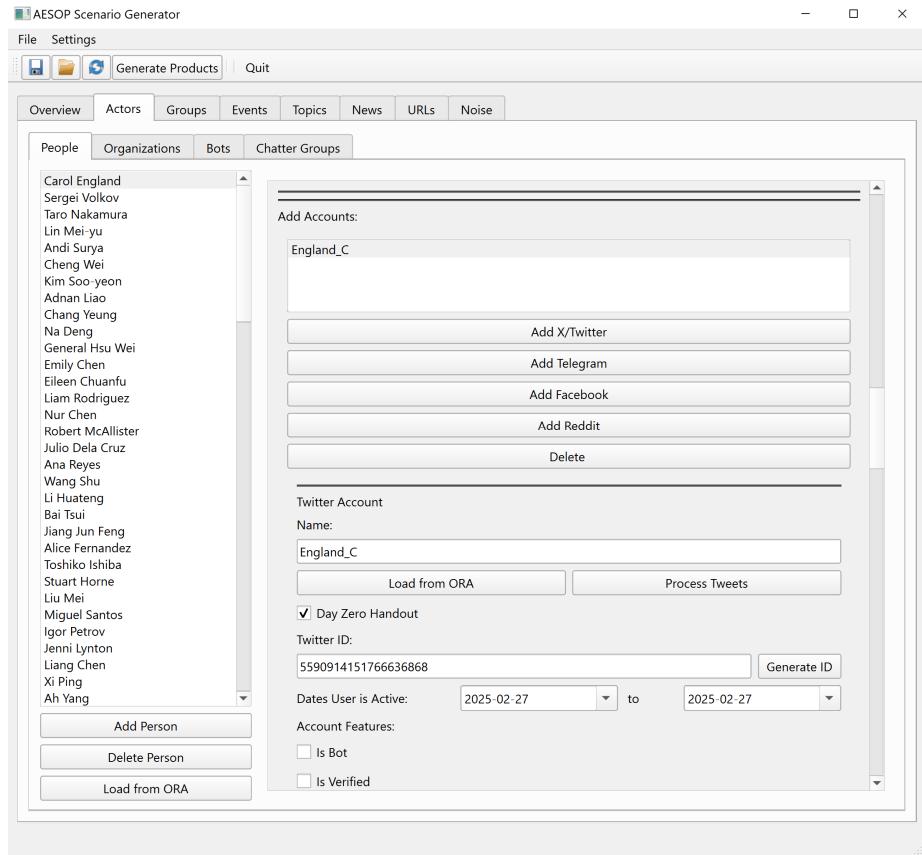


Figure 5.5: AESOP GUI for Account creation.

## Groups

Groups are how a planner organizes actors (not accounts). Actors are a part of groups in one of three ways - full member, source-only, or leader. If they are a full member, then they espouse the values and conduct actions in accordance with that group. If they are source-only, then full members of the group will repost, quote, or reply these source-only actors but the source-only actors themselves do not necessarily espouse the narratives of or act in accordance with that group. Actors can also be leaders of group - this distinction can only be given to actors that are also full-members and is to mark those actors that are the leaders of a group. Actors can be members of any number of groups and both full members and source only can be other groups - allowing for recursive placement for actors.

Groups also have a Group Topic Map (GTM) that is equivalent to the Actor Topic Map - a vector with a length equal to the number of topics and an opinion value from -5 to 5 for each

topic. This represents the initial position of the group on each topic. Associated with the GTM is the group's Topic-to-Topic Map (TTM). This is a matrix of topics against topics (defined later in this chapter) with values ranging from -1 to 1 that represent the conflated relationship between two topics for this group. This matrix helps synthetic generation entities understand the relationship between topics within this group.

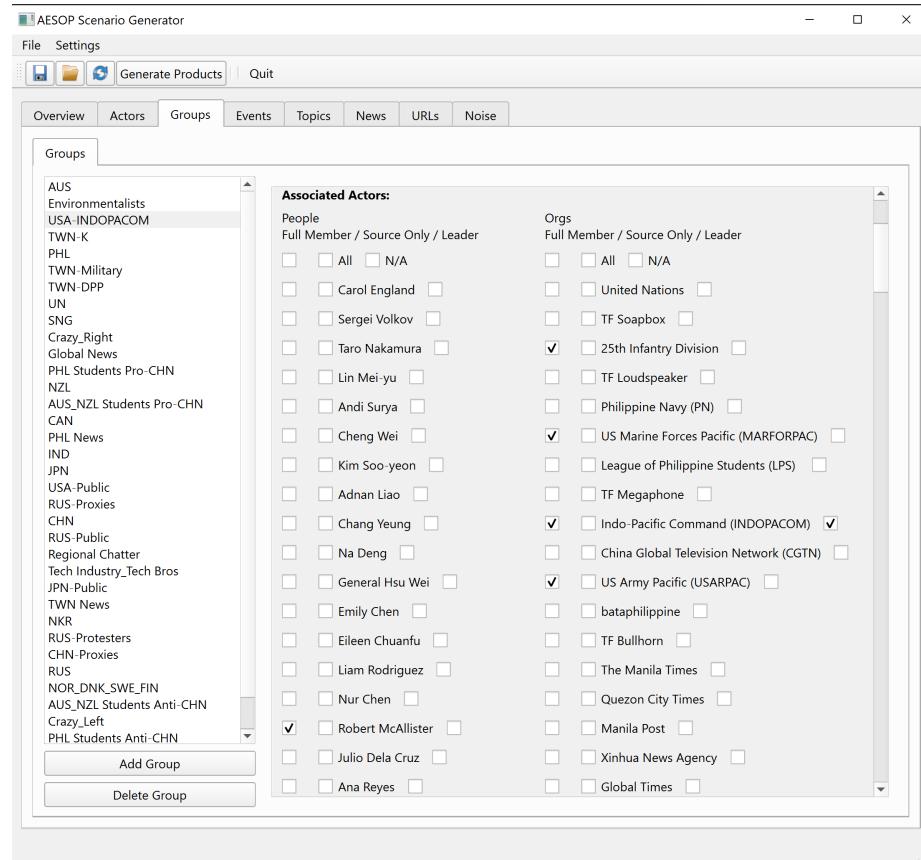


Figure 5.6: AESOP GUI for Group creation.

## Events

The WHAT category includes Events, Topics, News, and URLs. Events are the simplest to understand. Planners require a way to dictate what happens during the course of an exercise - these are called events. Each event is generally a purposeful attempt by the planner to solicit action on the part of the training audience.

### Event

- Event name
- Type [NATO event, Health Crisis, Election, Climate Event, Conflict/War, Military Event, Diplomatic Event, Other]
- Excitement Level

- Event Start
- Event End
- Other countries involved
- Regions/Areas involved
- Event Leader information
- Positive Hashtags
- Negative Hashtags
- Event Purpose/Summary
- Event Description
- Event Image

Upon selection of the event type, a large number of additional fields are available for input. As an example these are the additional fields for a military event:

#### **Military Event**

- Lead nation
- Involved services
- Lead service
- Leaders rank
- Other involved countries
- Live Fire Exercise

In general, planners are more concerned with the name of the event, the start and end times, the purpose of the event, and the level of excitement. Almost everything else can be auto-filled by the large-language model. The excitement level in particular is important. Ranging from 1-10, this value acts as a cue to the synthetic generator for how impactful this event should be within the generated data - determining how often this event is discussed in comparison to other concurrent events and if actors/accounts should take more actions than usual while this event is occurring.

Events are also where planners can specify additional injects into the exercise that are related to that event. There are four major types of inject - Intelligence Summaries, Fragmentary Orders, Press Releases, and Other. An Intelligence Summary (INTSUM) is developed as an inject when the planner wants to provide additional information along with an artifact to the training audience during the exercise scenario. For instance, if the training audience requests additional information about an account from the exercise controller, then the account information might be furnished to the training audience along with a corresponding INTSUM. A Fragmentary Order (FRAGO) is developed as an inject when the planner wants to forcibly redirect the training audience's attention or efforts. This is most often done to ensure the training audience conducts actions in accordance with a training objective. Within the US military a FRAGO is an amendment to an existing Operations Order; however, the concept can be applied to any incremental change of mission - even outside of military exercises. A planner develops a Press Release in order to help draw the attention of the training audience to an important event in a more subtle

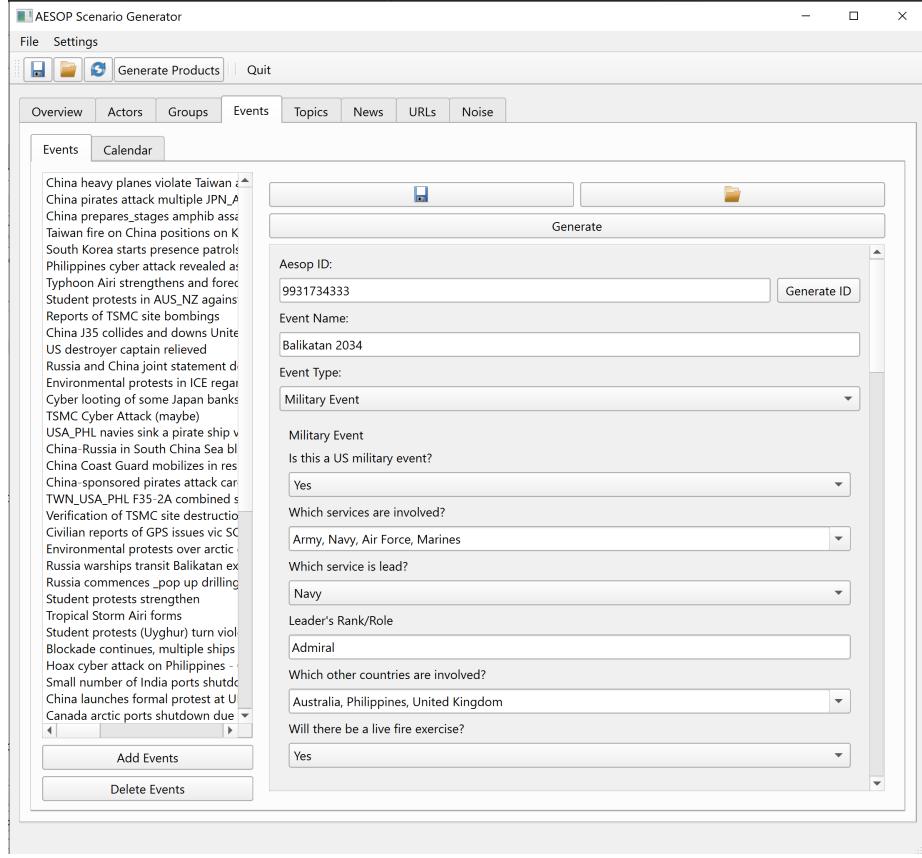


Figure 5.7: AESOP GUI for Event creation.

way - they are exactly what they sound like - a Press Release from a specified entity that discusses recent events. The Other type inject allows for the planner to freely interact with the LLM and develop whatever event-related product they might want to pass to the training audience - including more conventional cyberspace maneuvers. Surprisingly, both the OpenAI models and locally run LLM display a very good understanding of INTSUMs, FRAGOs, and Press Releases and little prompt engineering is required of the planner except to click a button adding the inject and requesting the LLM to fill out.

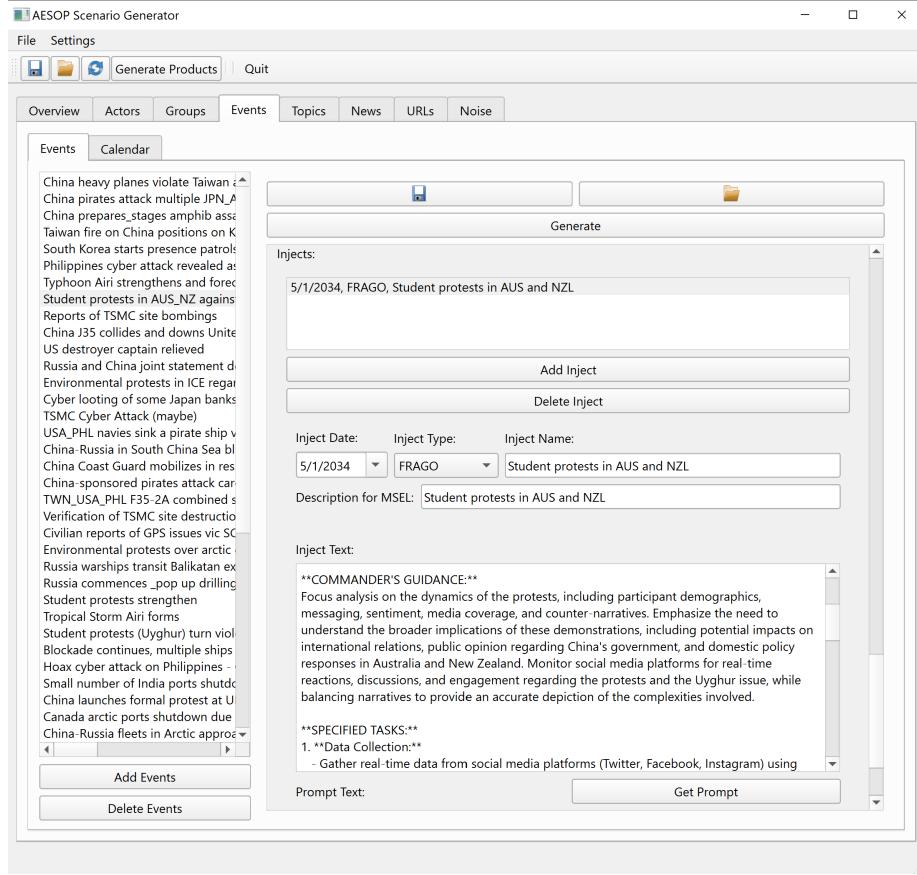


Figure 5.8: AESOP GUI for Inject creation.

## Topics

Topics are high-level opinion statements that are relevant to the scenario or the training objectives of the training audience. Example topics are: Balikatan is good for the Philippines, US intervention in the South China Sea has been good for the region, US anti-piracy actions have been ineffective and dangerous. Topics themselves have relatively few attributes.

### Topic

- Topic name
- Associated hashtags
- Topic description

However, topics are where narratives are added to the scenario. Narratives are added beneath each topic. In this case, a narrative is defined as messaging or discourse that reflects a particular stance on a topic. For instance, if the topic is "Dogs make the best pets" - a pro-narrative might be "Dogs are the most loyal of all domesticated animals" and an anti-narrative might be "Dogs chew everything and are destructive pets" - they each represent a stance on the topic as well as a distinct message. Narratives are where planners will spend most of their time. They need

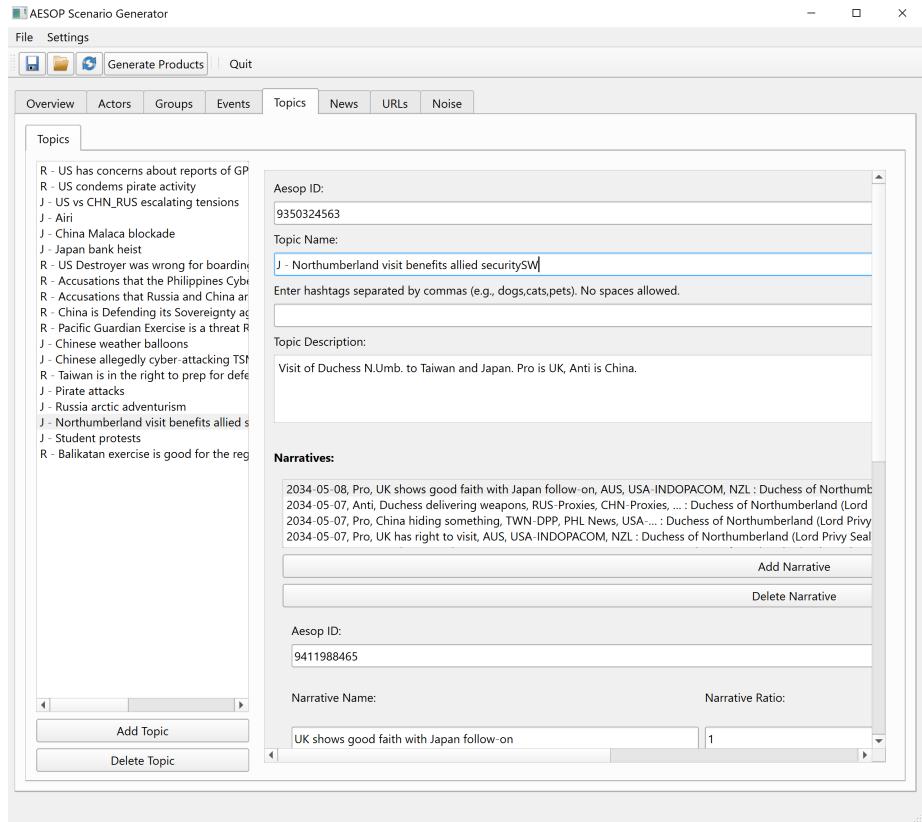


Figure 5.9: AESOP GUI for Topic creation.

to select topics that are required by the scenario and make sense to the training audience and then develop pro- and anti-topic narratives that reflect the stances of various NPCs within the scenario. Most often there are multiple pro- and anti-narratives for every topic that reflect the disparate groups engaging with each other in the scenario.

### Narrative

- Narrative Name
- Narrative Ratio
- Topic Stance [Pro, Anti, Neutral]
- Associated Groups
- Associated Events
- Hashtags
- Narrative Description
- Example Messages

Narratives are the bridge between the WHO (groups, actors, accounts) and the WHAT (events, news, URLs). Each narrative is associated with one or more groups that espouse it and also zero or more events that are linked to that narrative. A group linked to a narrative means that mem-

bers of that group will propagate that narrative. An event linked to a narrative means ties that event's excitement level to the narrative - influencing how often group members use that narrative through time.

Therefore, while the topics themselves are firmly in the WHAT category, these narratives encompass the HOW. Indeed, each narrative also includes a set of BEND maneuver preferences expressed as ratios that set the likelihood of those maneuvers being a part of that narrative's messaging.

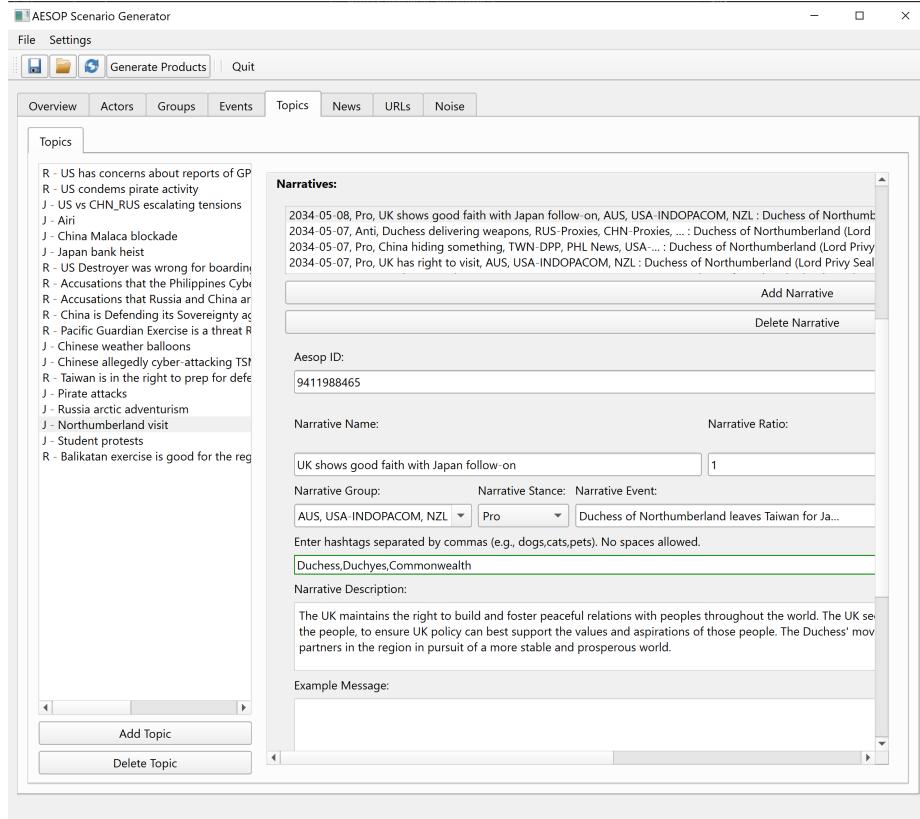


Figure 5.10: AESOP GUI for Narrative creation.

## News

In general, planners have now constructed who will be taking about what and how - but there are more supporting details required for robust and immersive training. In particular, messages rarely exist without corresponding news articles, images, memes, or other multimedia. AESOP provides planners the ability to create and connect news articles into the scenario. AESOP's News tab is where planners create news agencies, their websites, and these corresponding articles. These articles provide both immersion and additional avenues for information transfer to the training audience.

### News Agency

- Agency Name
- Agency Type [Real News, Pink Slime, Disinformation, Other]
- Available date range
- Editor
- Home country
- Targeted regions
- Bias [Extreme Left, Left, Center Left, Center, Center Right, Right, Extreme Right]
- Credibility [Low, Medium, High]
- Questionable Characteristics \*[Conspiracy Theories, Pseudoscience, Propaganda, Poor Sources, Failed Fact Checks]
- Summary
- History

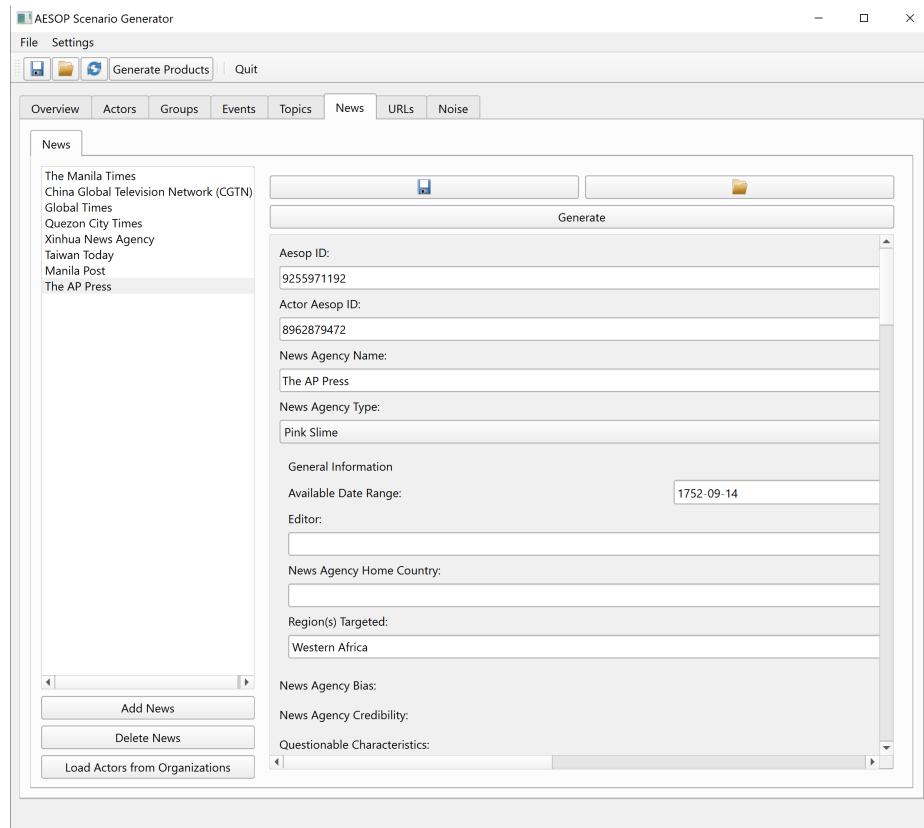


Figure 5.11: AESOP GUI for News Agency creation.

Individual articles have the following properties:

### **Article**

- Media Type [Real News, Pink Slime, Disinformation, Other]

- Article Topic Category [Industry, Health, Political, Ideas, Social, Lifestyle, Cultural, Awards, Education, Community, Other]
- Publication Date
- Headline
- URL
- Associated Narratives
- Author
- Targeted Region
- Article Bias [Extreme Left, Left, Center Left, Center, Center Right, Right, Extreme Right]
- Article Credibility [Low, Medium, High]
- Agency Type [Real News, Pink Slime, Disinformation, Other]
- Available date range
- Editor
- Home country
- Targeted regions
- Bias [Extreme Left, Left, Center Left, Center, Center Right, Right, Extreme Right]
- Credibility [Low, Medium, High]
- Questionable Characteristics \*[Conspiracy Theories, Pseudoscience, Propaganda, Poor Sources, Failed Fact Checks]
- Number of Paragraphs
- Summary

AESOP supports the individual creation of both news agencies and their associated articles. However, if planners have already created an Actor that is of the type News Agency, then AESOP provides a push button capability to auto-create a matching news agency for every Actor of that type. Additionally, if the planner has already completed the Groups, Events, and Topics, then AESOP can auto-create articles for every narrative and event combination corresponding to the news agency actor's group membership. In this way, large numbers of appropriate articles can be quickly created to support the scenario without intervention from the planner.

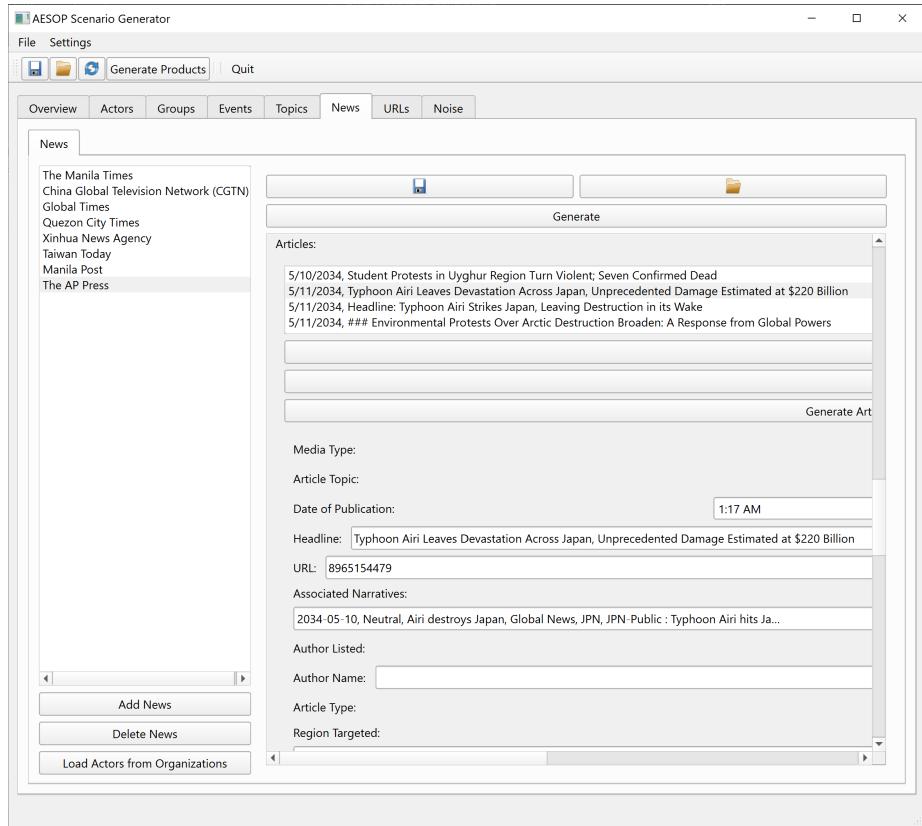


Figure 5.12: AESOP GUI for Article creation.

## URLs

Finally, planners need to provide URLs for all media that can be referenced by NPCs within the exercise scenario. For news articles, AESOP will auto-populate a URL for each of the news articles. However, for externally created videos, images, and other multimedia, planners will need to create URL entries, choose the dates when the media is available for reference, and decide what narratives are associated with these media. AESOP does not host any of this media.

### URL

- URL
- Date Available
- Type [SITE, IMAGE, VIDEO, OTHER]
- Associated Narratives
- URL description

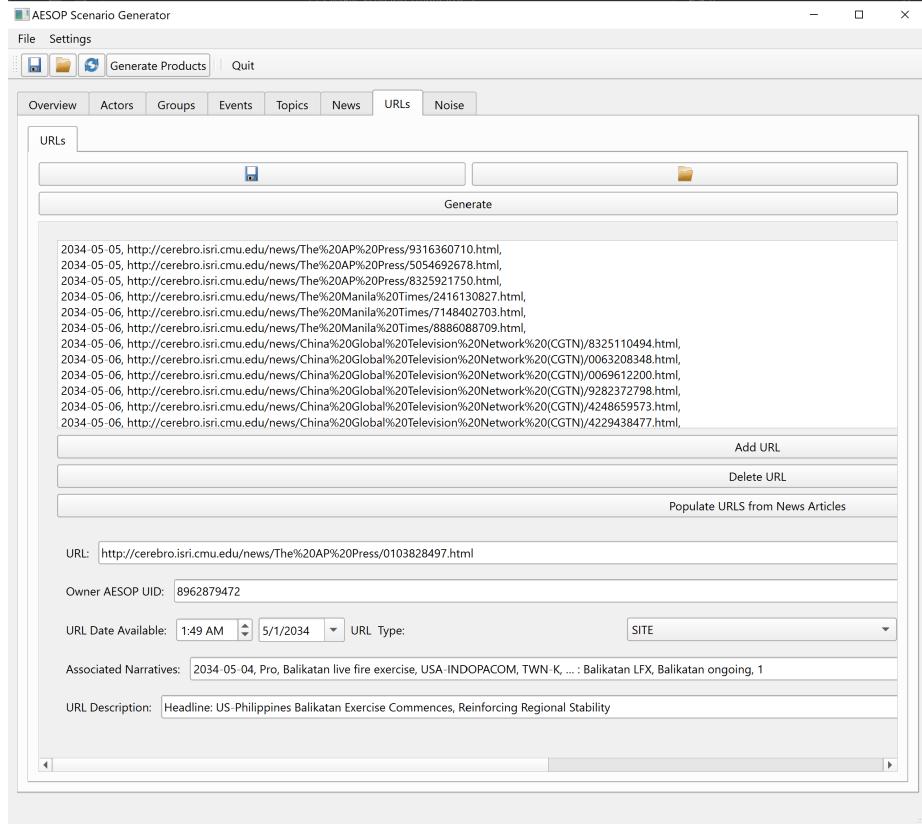


Figure 5.13: AESOP GUI for URL creation.

## Noise

Thus far, planners have concerned themselves with what can be thought of as "the needles" - these are what planners want the training audience to respond to. AESOP's noise tab is where the planners define the "the haystack" - the background noise from which the training audience needs to glean the needles. In its current state, this tab is primarily concerned with X/Twitter noise. Planners need to determine the correct number of overall messages for the haystack, where these messages come from, and the date range for the messages. Additionally, because AESOP is trying to facilitate realistic training, it is important that planners consider what imaginary search terms were used to pull from the X API to get the needle + haystack dataset. This lets synthetic generators know what topics should be present in the dataset and prevents completely spurious data from being present.

### Noise

- Total messages
- Noise locations
- Date range
- Search terms
- Additional notes

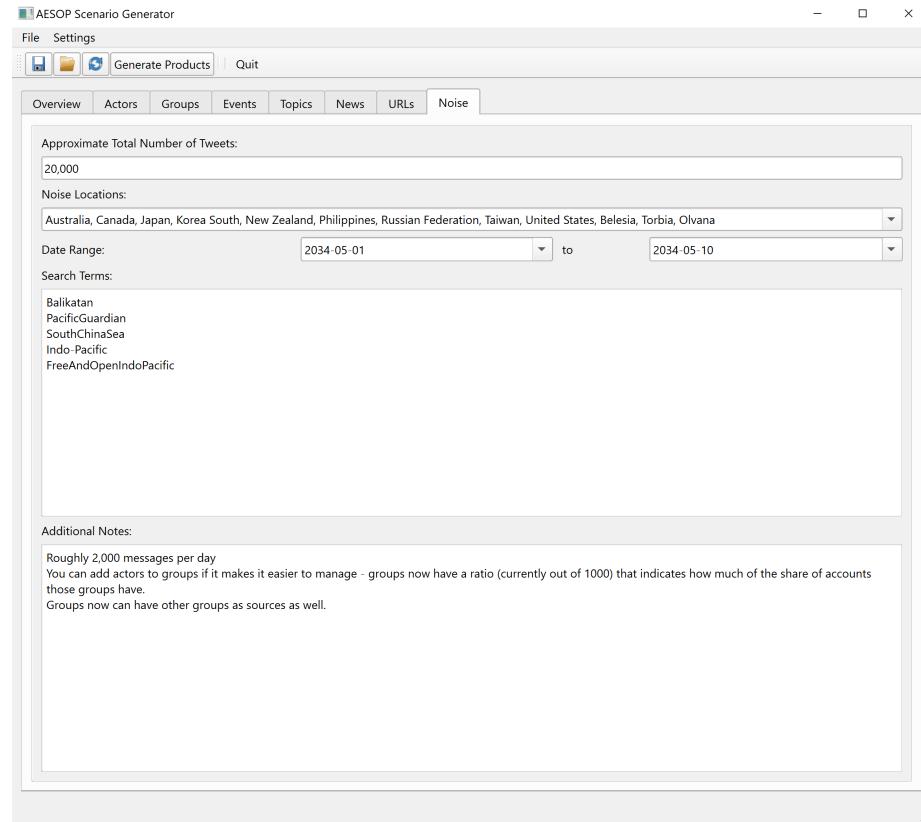


Figure 5.14: AESOP GUI for Noise creation.

## 5.4 AESOP Outputs

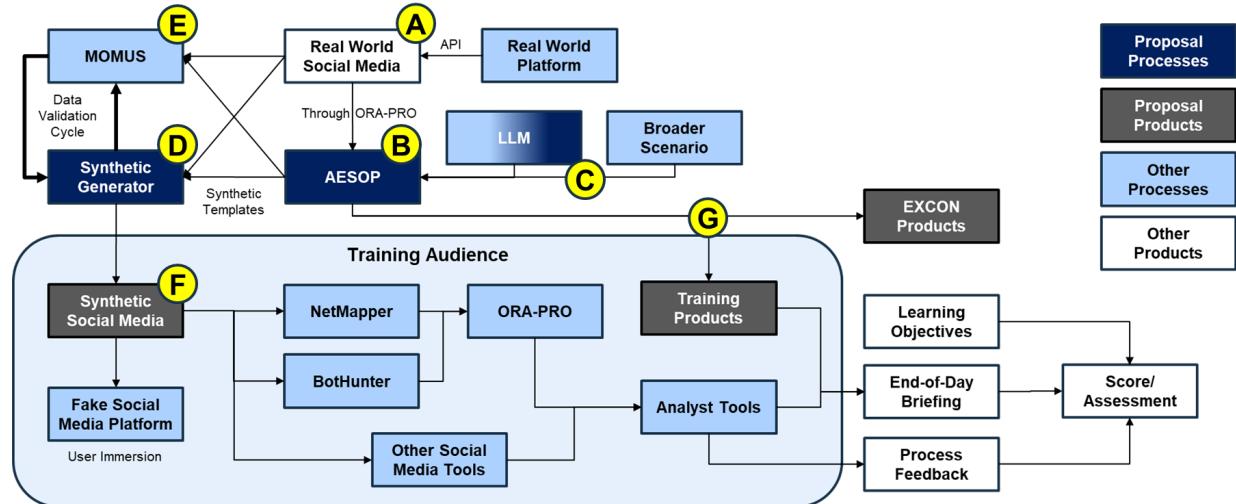


Figure 5.15: Project OMEN training flow.

After completing scenario construction within AESOP (B in Fig. 5.15), planners use AESOP to generate two main components. They generate synthetic templates for consumption by synthetic generators (D in Fig. 5.15) and generate products for the training audience and the exercise controllers (G in Fig. 5.15). These are two very different products. The products for the training audience and exercise controllers are meant to be artifacts for inclusion in the actual exercise. As such, they come in formats that should be familiar: MS Word Documents and MS Excel spreadsheets. As a bonus, AESOP can apply cover pages, security release footers, and custom formats to all of these "paper" products. These documents are output in separate folders designated for either the training audience or exercise control. Examples of these products can be found in Appendix ??

Standard (Acronym)	Purpose	Key Components	Primary Use Case	Data Format	Adoption	Website Reference
Structured Threat Information Expression (STIX)	Cyber threat intelligence, tracking actors, events, and campaigns	Identity, Threat Actor, Campaign, Incident, Observables	Tracking cyber threats, disinformation campaigns, and influence operations	JSON, XML	Cybersecurity, threat intelligence, law enforcement	<a href="https://oasis-open.github.io/cti-documentation/">https://oasis-open.github.io/cti-documentation/</a>
Activity Streams 2.0 (AS2)	Social media activity representation	Actor, Object, Verb, Target, Group	Modeling social media interactions and behavioral dynamics	JSON-LD, ActivityStreams format	Decentralized social media platforms and Web3 applications	<a href="https://www.w3.org/TR/activitystreams-core/">https://www.w3.org/TR/activitystreams-core/</a>
Friend of a Friend (FOAF)	Describes social relationships and online identities	Person, Group, OnlineAccount, Knows	Semantic web applications and social network analysis	RDF (Resource Description Framework)	Semantic web, academic research, social network modeling	<a href="http://xmlns.com/foaf/spec/">http://xmlns.com/foaf/spec/</a>
Semantically-Interlinked Online Communities (SIOC)	Defines social media discussions and interactions	UserAccount, Post, Forum, Community, Topic	Online community tracking and engagement analysis	RDF, XML	Online discussion forums, enterprise knowledge management	<a href="https://web.archive.org/web/20220331224416/http://sioc-project.org/">https://web.archive.org/web/20220331224416/http://sioc-project.org/</a> <a href="https://www.dni.gov/index.php/who-we-are/organizations/ic-cio/ic-technical-specifications/information-resource-metadata">https://www.dni.gov/index.php/who-we-are/organizations/ic-cio/ic-technical-specifications/information-resource-metadata</a>
Intelligence Community Information Resource Metadata (IC-IRM)	Intelligence metadata for information security, discovery, and analysis	Resource descriptions, classification, access control, XML/HTML encoding	Intelligence and national security data management	XML, HTML metadata(IC)	US Intelligence Community	
Unnamed Synthetic Social Media Scenario Data Standard (USS-MSDS)	Describes actors, events, groups, and narratives within a scenario	Actor, Account, Event, Group, Topic, Narrative, Article, URL	Communicating the structure and component of a social media scenario from the scenario planner to a synthetic generator	JSON	NONE	NONE

Figure 5.16: A comparison of related data standards.

The synthetic templates are not meant for the training audience or the exercise controllers, instead they are meant to define the scenario for the synthetic generators. To facilitate the transition of the planners intent as expressed in AESOP to the information required for the generation of that intent, I propose a data standard for synthetic social media scenarios. In general, it is a very bad idea to create new data standards.[56] However, there are no existing candidates within this space. Structure Threat Information Expression (STIX) has actor and event enumeration - similar to AESOP's outputs; however, it focuses on cyber threat intelligence and the fields required for creating social media from the actors and events are missing.[59]. Activity Streams 2.0 also represents social media activity using an Actor format. Unfortunately, its focus is on providing a method for itemizing actual behavior rather than quantifying typical behavioral patterns needed for synthetic generation.[66] Friend of a Friend (FOAF) allows for the compact representation of an existing social network but it does not provide sufficient information for constructing or deriving new ones.[19]. The Semantically-Interlinked Online Communities (SIOC) standard is closer to what is required for synthetic generation but lacks definitions for the reasons (Events, Narratives) behind actions taken by its equivalent Account entity - UserAccount.[18] Lastly, the US Intelligence Community has a robust standard for metadata - the Intelligence Community - Information Resource Metadata (IC-IRM) standard. This standard in no way describes social media network components for use in synthetic generation. However, its broad adoption within the DoD, its extensibility, continuous development, and ever broadening scope mean that almost

all burgeoning data standards are compared to it.[60] The IC-IRM does not currently have any overlap with required components for synthetic data generation. The complete data standard used for the synthetic templates can be found in Appendix C.

A full list of AESOP outputs can be found in Table 5.2. The archive files are the save system that AESOP uses, the synthetic templates are set against the data standard, the EXCON column are the products furnished to the exercise control and the Participants column is for the training audience.

## Safety and Ethics

In a Joint Cybersecurity Advisory released by the Federal Bureau of Investigation (FBI), the Cyber National Mission Force (CMNF), the Netherlands General Intelligence and Security Service (AIVD), Netherlands Military Intelligence and Security Service (MIVD), the Netherlands Police (DNP), and the Canadian Center for Cyber Security (CCCS), they accused Russian affiliated groups of using sophisticated AI-enhanced software packages to perform information maneuvers online. [36] The software package was called Meliorator - with a front-end called Brigadir and a back-end called Taras. Brigadir has tabs for "souls" - similar to AESOP's Actors - and "thoughts" - roughly analogous to AESOP's narratives. There exists some level of superficial similarity between Meliorator and AESOP. However, the Meliorator output through Taras includes code that creates and manipulates actual online accounts on live social media platforms in concert with fake personas. AESOP only outputs the characteristics of actors. There is no connection to any real world platform - no scripting, no scraping, no interface whatsoever. The purpose of AESOP is to enable the construction of training scenarios so that real-world interaction is not unnecessary.

## 5.5 Future Work

AESOP is an effective tool for quickly putting together a high-quality scenario. However, it is primarily an engineering solution with a simplified GUI skin. There are significant improvements that should be made in the future that would increase its functionality and would also increase planner's quality of life when interacting with AESOP. As priorities, the following should be accomplished first:

AESOP should be tied with retrieval-augmented generation (RAG) to limit hallucinations and provide additional support to the mad-lib method of prompt-engineering used to scaffold the planner. RAG would allow planners to rely upon a corpus of validated material to supplement the LLMs trained knowledge.[53]

AESOP should integrate the ability to populate personas drawn from population modeling of a demographic. This would provide an effective method for determining the types of personas that are representative of a demographic, rather than relying upon the planner to select

AESOP should allow for the quick integration of multiple LLMs for specialized purposes. Currently, the LLM is set at the global level within AESOP and this pushes AESOP toward using large generalized models - favoring commercial models over local LLMs. The ability for AESOP to operate with purely local LLMs does already exist but specifying multiple, smaller,

specialized LLMs for agent creation versus journal/article writing versus press releases would increase quality while still allowing for disconnected operation.

## 5.6 Conclusions

AESOP provides planner's a faster way of integrating social media scenarios into an existing exercise or training event. It also helps them plan independent social media exercises without high-levels of domain knowledge. This two fold reduction in effort - both in time and in requisite domain knowledge - represents a major step forward in social media exercise planning. Additionally, the use of a data standard means that as AESOP continues to grow and change - or is replaced by other planning tools - the connective tissue between the planner and the generation capability remains unsevered.

Table 5.2: AESOP Outputs and File Types

Feature	Archive File	Synthetic Template	EXCON	Participants
<b>Agents</b>				
People	JSON	JSON	DOCX	DOCX
Organizations	JSON	JSON	DOCX	DOCX
Bots	JSON	JSON	DOCX	DOCX
<b>Telegram</b>				
Account	JSON	JSON		
Channel	JSON	JSON		
Messages	JSON	JSON		
<b>X/Twitter</b>				
Account	JSON	JSON		
Messages	JSON	JSON		
<b>Groups</b>				
Group	JSON	JSON	DOCX	
<b>Topics</b>				
Topic	JSON	JSON	DOCX	
Narrative	JSON	JSON	DOCX	
<b>Events</b>				
Event Summary	JSON	JSON	DOCX	DOCX
Fragmentory Orders	JSON		DOCX	DOCX
Press Releases	JSON		DOCX	DOCX
Intelligence Reports	JSON		DOCX	DOCX
Other	JSON		DOCX	DOCX
<b>News</b>				
News Agency	JSON			
News Articles	JSON			
<b>URLs</b>				
URL	JSON	JSON		
<b>Master Synch Event List</b>				
Populated MSEL	JSON		XLSX	
<b>Scenario Overview</b>				
Scenario Description	JSON		DOCX	DOCX
Mission	JSON		DOCX	DOCX
Higher HQ Mission	JSON		DOCX	DOCX
Commander's Guidance	JSON		DOCX	DOCX
Strategic Coms Guidance	JSON		DOCX	DOCX

# Chapter 6

## Synthetic Social Media Creation

### 6.1 Research Questions

The creation of realistic, dynamic, and controllable synthetic social media data is a critical component of operationalizing social-cybersecurity training and evaluation. While real-world data provides valuable context, it often lacks the adaptability, interactivity, and narrative specificity needed for focused training objectives—particularly when the goal is to understand and identify complex influence maneuvers.

**The key research questions for this chapter are:**

- How can we create synthetic data that includes BEND maneuvers to support a training scenario?
- How can we appropriately leverage large-language models in the creation of synthetic social media datasets?

To answer these questions, I describe the design and implementation of a hybrid simulation framework built to generate synthetic social media corpora. This framework integrates the structure and intent derived from training scenarios authored in AESOP with the flexibility and realism of large language models (LLMs), yielding datasets that reflect both network-level interactions and narrative maneuvering.

The chapter begins by surveying the broader landscape of synthetic data generation, outlining the strengths and limitations of both top-down (system-level) and bottom-up (agent-based) approaches. I then introduce SynTel and SynX, the agent-based generators developed for this thesis, which combine traditional simulation logic with LLM-powered message construction. By controlling when and how LLMs are used—specifically for generating realistic text rather than building the entire network—I sidestep key scalability challenges while maximizing narrative fidelity.

Finally, I walk through the end-to-end process of how SynX operates: from determining agent actions to generating messages consistent with BEND maneuvers and validating them using effects-based detection methods. This approach ensures that the resulting synthetic data is

both analytically useful and operationally relevant for training, experimentation, and scenario-based planning.

## 6.2 Synthetic Generation Approaches

There are two main approaches to synthetic generation, top-down and bottom-up. These are also referred to as macro-level or system-based and micro-level.[55] [25] In the top-down approach, a simulation assumes a desired heterogeneous/multi-modal network fabric based on real data and then fills that fabric with appropriate actors, message types, and narratives that match nodes and link types. Alternatively, using the bottom-up approach, the simulation starts with agents programmed from first principles with detailed interaction rule sets. The agents then interact with each other hoping for emergent networks and narratives that are realistic and relevant.[?] This dichotomy mirrors approaches by others, including Chang et al. in 2024.[24] Interestingly, Chang et al. also point out a new dichotomy - the use of large-language models versus traditional network construction algorithms. They explore an LLM-only methodology that revealed the strength of the bottom-up approach but also found limitations with their LLM-only system. That is, their LLM-only methodology is not capable of scaling when combined with the bottom-up approach. The LLM prompts require iterating through all personas with each persona provided the information about all other personas. This is unwieldy, even if there are only dozens of actors in a network.

For SynTel/X and its synthetic generation of social media networks, I opted for a hybrid approach along both spectra. SynTel/X is grounded in agent-based simulation, with each Actor/Account acting independently based upon an action rule-set derived from their attributes and features. However, rather than let agents organically build their own networks and form narratives, I dictate who their groups members are and what narratives they can express. This approach might seem obvious from the data standard composition of the synthetic templates supplied by AESOP, but the data standards were derived post hoc from the system. A data standard should not drive how a system executes.

Furthermore, rather than using an LLM-only system, SynTel/X incorporates LLMs only in the final stages of generation. Network construction is done entirely by traditional algorithms, and LLMs are leveraged only for narrative construction - this eliminates the need for prompts to have information about every other actor, solving scalability issues, and also leverages LLMs for what they are best at - sounding like real people.[39] [64]

This hybrid approach reflects a difference in the desired outcome compared to other efforts. Generally, other approaches are looking to maximize one of three outcomes: the re-creation of the structure of real social media,[24] the recreation of human social media text output,[69] or the generation of a specific set of synthetic training data.[71][38][51] The SynTel/X output needs to accomplish the first two, but can eschew the latter. In order to provide realistic training, SynTel/X needs to produce social media data equivalent to the social media data collected directly from an API of a real-world platform.

There are three related simulations from which SynX draws:

The twitter\_sim2.0 model – as outlined by Blane, Moffit and Carley in 2021 [15] – is a model focused on Twitter interactions. The simulation accounts for both emotion and logic – ensuring

Table 6.1: Docking Lite

Feature	SynX	twitter_sim2.0Construct	LLM-Social-Network
Media Agents	✓	✓	✓
Opinion Leaders	✓	✓	✓
Information Access	✓	✓	✓
General Memory	✓	✓	✓
Homophily	✓	✓	✓
Limited Attention	✓	✓	✓
Dynamic Network	✓	✓	✓
Emotional Response		✓	✓
All BEND Maneuvers	✓		✓
Full Diffusion			✓
Live Visualization			✓
AESOP inputs	✓		
Full Messages	✓		
Context Inputs	✓		✓
X APIv1/v2 Format	✓		

tweets that emotionally correspond with a recipient have magnified effects. Importantly, it cannot take in existing or user generated tweets as context and while it includes BEND maneuvers it is not interpretable by ORA-Pro. It also does not produce a message corpus for external evaluation.

Construct is a simulation framework for implementing agent-based modeling in C++20[34]. Construct can parse DynetML files from ORA or CSV. It allows for the custom creation of models - including those for information diffusion. However, it does not produce a message corpus.

LLM-Social-Network by Chang et al. in 2024, leverages large language models to reconstruct agent x agent networks.[24] It also stops short of generating an associated message corpus and also offloads agent-based decision making to an LLM.

Table 6.1 shows a comparison between these simulations with respect to their nodes and features of network construction - it does not take into account their simulation processes.

## 6.3 SynX

SynX and SynTel are synthetic generators that take synthetic templates from AESOP (required), existing X/Telegram corpora (optional), and injected tweets/messages from the training audience (optional) and use an LLM to output synthetic social media data. SynX produces X/Twitter APIv1 data.

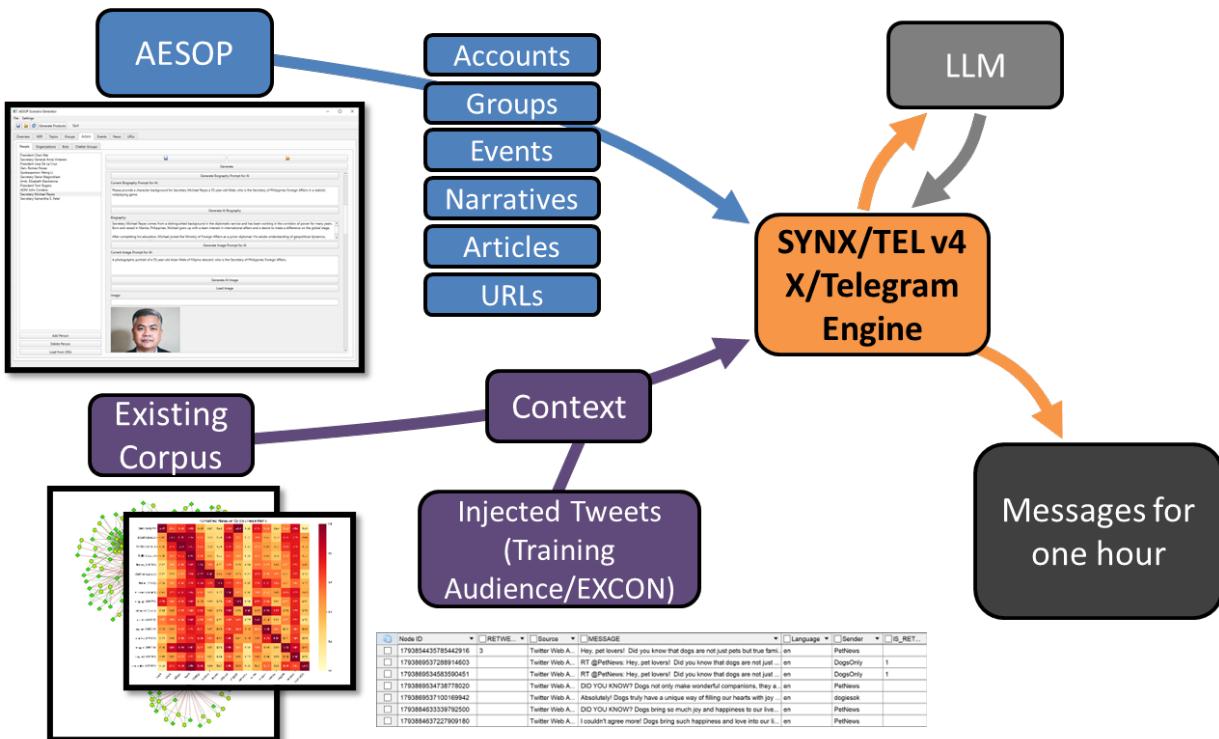


Figure 6.1: Overview diagram of SynTel/X

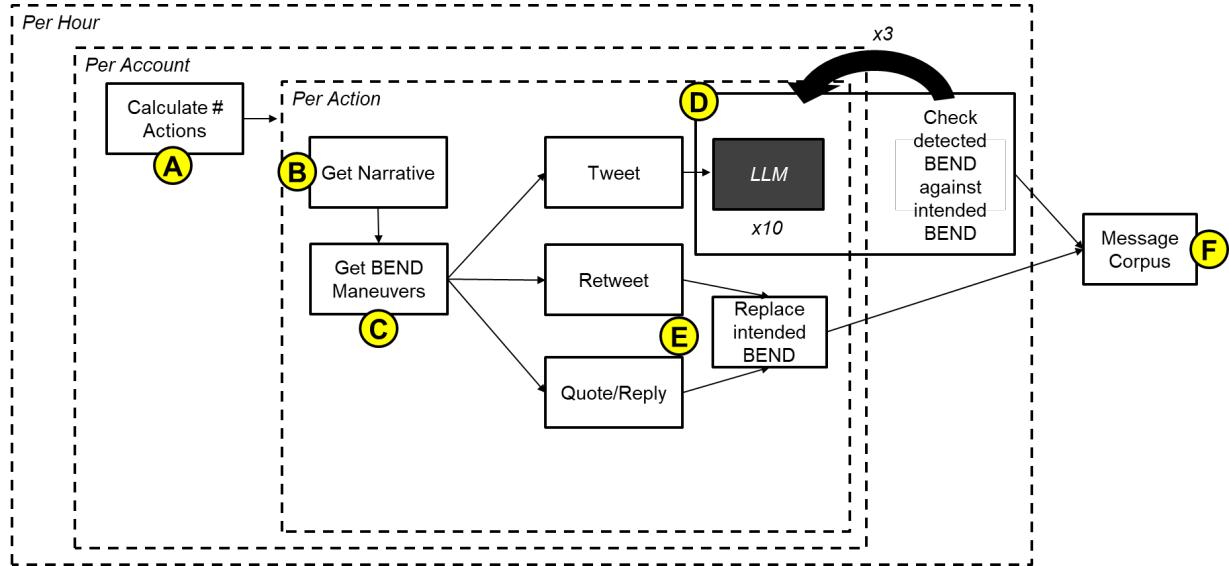


Figure 6.2: Example SynX logic flow.

## SynX Flow Overview

- A) Calculate Actions: Per hour, per account, SynX determines the number of actions to be taken by the account.
- B) Find a Narrative: For each action, SynX determines a narrative from a weighted sampling of all narratives associated with groups of which the Actor/Account is a member.
- C) Determine BEND Maneuvers: Based on the BEND maneuvers associated with that narrative and conflated with the BEND maneuvers associated with the Actor/Account, a seed BEND maneuver is selected and then a full chain of maneuvers is determined.
- D) Construct the Prompt: If the BEND maneuvers suggest an original message, then a message shell is created and metadata adjusted to match the intended BEND maneuvers. Then a prompt is constructed and an LLM call is made to create a message - this may happen several times as SynX attempts to make sure that the intended BEND maneuvers are detected within the output message.
- E) Find Another Message: If the BEND maneuvers suggest a derivative message, then an original message is selected from this narrative, and the derivative message is constructed with a partial replacement of the BEND maneuvers.
- F) Produce Output: All of the messages are combined into the message corpus.

### A) Calculate Actions

SynX simulates an hour of X messages at a time. To do this, all X accounts go through the decision-flow process summarized above. First, the account determines how many actions to

take in a given hour.

Let:

- $R_o = [r_{o,\min}, r_{o,\max}]$  be the range of original tweets per day
- $R_r = [r_{r,\min}, r_{r,\max}]$  be the range of retweets per day
- $R_q = [r_{q,\min}, r_{q,\max}]$  be the range of quote/reply tweets per day
- $E$  be the excite number for the day (the max of all excite variables from scenario events occurring during that day)

Then:

$$\begin{aligned}\text{min\_posts} &= r_{o,\min} + r_{r,\min} + r_{q,\min} \\ \text{max\_posts} &= r_{o,\max} + r_{r,\max} + r_{q,\max} \\ \text{adjusted\_min} &= \lfloor \text{min\_posts} \times E \rfloor \\ \text{adjusted\_max} &= \lfloor \text{max\_posts} \times E \rfloor\end{aligned}$$

Finally, the number of posts to generate is randomly selected from the integer interval:

$$\text{post\_count} \sim \mathcal{U}(\text{adjusted\_min}, \text{adjusted\_max})$$

However, the number of posts per day is not sufficient for SynX, because the accounts are run per hour. In order to transform the total posts per day into a probability of a post(s) occurring during a single hour, we also need the active daily schedule of an account - this is given in the synthetic template for that account.

Let  $h \in \{0, 1, \dots, 23\}$  be the hour of the day, and let  $T$  be the total number of posts the account will make in a day. If the active schedule of the account is from 0900 to 1800 then we want higher probabilities during that time and reduced but tapering probabilities during other times.

We can therefore define the unnormalized hourly probability  $P(h)$  as:

$$P(h) = \begin{cases} 1 & \text{if } 9 \leq h < 19 \quad (\text{flat period}) \\ 1 - \frac{r(h)}{11} & \text{if } h \in \{18, 19, \dots, 23, 0, 1, 2, 3, 4\} \quad (\text{tapering}) \\ \frac{s(h)}{5} & \text{if } 5 \leq h < 9 \quad (\text{rising}) \\ 0 & \text{otherwise} \end{cases}$$

Where:

- $r(h)$  is the rank (0-indexed) of hour  $h$  in the tapering list:  $[18, 19, 20, 21, 22, 23, 0, 1, 2, 3, 4]$
- $s(h)$  is the rank (0-indexed) of hour  $h$  in the rising list:  $[4, 5, 6, 7, 8]$

Normalize the probabilities:

$$\tilde{P}(h) = \frac{P(h)}{\sum_{i=0}^{23} P(i)}$$

Then the final expected number of posts at hour  $h$  is:

$$\text{ExpectedPosts}(h) = T \cdot \tilde{P}(h)$$

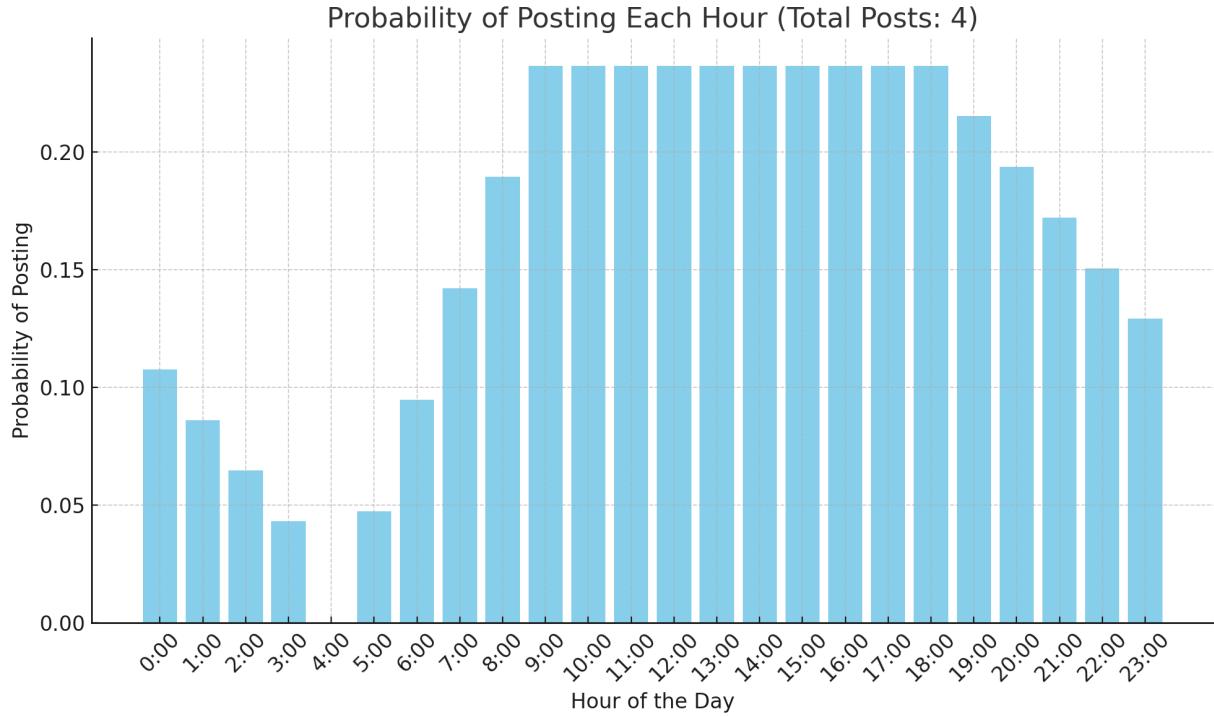


Figure 6.3: Graph of probability of posting each hour for a user active 0900 to 1800 who posts 4 times a day.

Fig. 6.3, graphically depicts the probability of posting each hour for a user with four total posts in a day. The distribution is flat from 09:00 to 18:00, tapers off between 18:00 and 04:00, and gradually increases from 04:00 to 09:00.

## B) Find a Narrative

Now that we have calculated the number of actions to be taken in a period, we need to know what type of actions will be taken. Naively, the original posts, retweets, quotes/replies would not have been amalgamated into total actions, and each would be run separately. However, while it makes sense for planners and even social media analysis tools to calculate individual ranges for each type of posting, this does not work well at an agent level. The outcome of the training is the analysis and understanding of the BEND maneuvers. Ostensibly, an account posts not because

of a number on its template but because that account is trying to execute a BEND maneuver. Because certain BEND maneuvers are closely associated with each type of X post, we must have accounts choose BEND maneuvers that determine a post type rather than choosing a post type that then dictates BEND maneuvers.

Therefore, for each action to be taken in a period, the account finds a narrative. This narrative is drawn as a weighted random selection or categorical distribution from all narratives associated with all groups of which the account's owning actor is a member.

Let:

- $A$  be an actor
- $\mathcal{G}(A) = \{G_1, G_2, \dots, G_m\}$  be the set of groups that actor  $A$  belongs to
- $\mathcal{N}(G_j)$  be the set of narratives associated with group  $G_j$
- $\mathcal{N}_A = \bigcup_{j=1}^m \mathcal{N}(G_j)$  be the full set of candidate narratives for actor  $A$
- Each narrative  $N_i \in \mathcal{N}_A$  has an associated weight  $w_i > 0$

Define the probability of selecting narrative  $N_i \in \mathcal{N}_A$  as:

$$P(N_i | A) = \frac{w_i}{\sum_{N_j \in \mathcal{N}_A} w_j}$$

Then, the selected narrative  $N^*$  for the message is drawn from the categorical distribution:

$$N^* \sim \text{Categorical}(\{P(N_i | A)\}_{N_i \in \mathcal{N}_A})$$

### C) Determine BEND Maneuvers

After a narrative is selected per action, a second weighted random selection is executed. The algorithm for this is similar to the narrative selection above except this time the algorithm is choosing BEND maneuvers based upon weights associated with that narrative conflated with the individual account BEND maneuver weights.

Let:

- $M = \{M_1, M_2, \dots, M_{16}\}$  be the set of BEND maneuvers
- $w_i^{(n)}$  be the weight of maneuver  $M_i$  from the selected narrative
- $w_i^{(a)}$  be the weight of maneuver  $M_i$  from the actor profile

Define the combined weight for each maneuver  $M_i$  as:

$$w_i^{(\text{combined})} = \left(w_i^{(n)}\right)^{2/3} \cdot \left(w_i^{(a)}\right)^{1/3}$$

Normalize to obtain a probability distribution over the 16 maneuvers:

$$P(M_i) = \frac{w_i^{(\text{combined})}}{\sum_{j=1}^{16} w_j^{(\text{combined})}}$$

Then sample one maneuver  $M^*$  from the categorical distribution:

$$M^* \sim \text{Categorical}(\{P(M_i)\}_{i=1}^{16})$$

From this seed BEND maneuver we can use a variable-order Markov model with set-based context and non-replacement sampling to determine the total BEND maneuvers for the action.

### BEND Sequence Generation from a Start Maneuver

Given a starting maneuver  $m_0$ , we build a sequence  $S = [m_0, m_1, \dots, m_k]$  by repeatedly sampling from the conditional probability distribution:

#### 1. Initial Maneuver

$$S_0 = [m_0]$$

#### 2. At Each Step

Let the current sequence be  $S_t = [m_0, m_1, \dots, m_t]$ , and let

$$\mathcal{S}_t = \text{sorted}(S_t)$$

Then, compute the conditional probabilities for the next maneuver  $m_{t+1}$ :

$$P(m_{t+1} | \mathcal{S}_t)$$

#### 3. Mask Already Chosen Maneuvers

To ensure uniqueness, set:

$$P(m | \mathcal{S}_t) = 0 \quad \text{if } m \in \mathcal{S}_t$$

#### 4. Normalize Probabilities

$$\hat{P}(m | \mathcal{S}_t) = \frac{P(m | \mathcal{S}_t)}{\sum_{m' \notin \mathcal{S}_t} P(m' | \mathcal{S}_t) + P(\text{END} | \mathcal{S}_t)}$$

#### 5. Sampling

Sample  $m_{t+1} \sim \hat{P}(\cdot | \mathcal{S}_t)$

## 6. Termination Condition

If  $m_{t+1} = \text{END}$ , stop.

From this we arrive at our full sequence of BEND maneuvers:  $S = [m_0, m_1, \dots, m_k]$ . This set of BEND maneuvers determines if this message is an original tweet, a retweet, a reply, or a quote. If the BEND maneuvers include back, engage, or neutralize, then the post is likely a derived type (retweet, reply, or a quote) as these BEND maneuvers are closely associated with these types of posts.

Let:

- $B = \text{BEND}(narrative)$

- $D = \text{isDerived}(B) = \begin{cases} 1 & \text{if } B_{\text{back}} = 1 \text{ or } B_{\text{engage}} = 1 \text{ or } B_{\text{neutralize}} = 1 \\ 0 & \text{otherwise} \end{cases}$

Then the tweet type  $T$  is sampled as follows:

$$T = \begin{cases} \text{DerivedType}() & \text{with probability } \frac{2}{3}, \text{ if } D = 1 \\ \text{Tweet} & \text{with probability } \frac{1}{3}, \text{ if } D = 1 \\ \text{Tweet} & \text{if } D = 0 \end{cases}$$

Where:

$$\text{DerivedType}() = \begin{cases} \text{Retweet} & \text{with probability } \frac{3}{5} \\ \text{Reply} & \text{with probability } \frac{1}{5} \\ \text{Quote} & \text{with probability } \frac{1}{5} \end{cases}$$

## D) Construct the Prompt

If the BEND maneuvers suggest an original message, then a message shell is created and metadata adjusted to match the intended BEND maneuvers. This is important because, as will be discussed at length in the following chapter, current methods of detecting BEND maneuvers operate on a per message level and evaluate for maneuvers based on both the content of the message and the metadata. As a general rule, metadata is more influential in determining network maneuvers, while the content of the message itself is more influential in determining narrative maneuvers. In particular, the term metadata is used here to describe anything outside of the message field (text or full\_text) of the tweet - to include portions of the message field that are enumerated externally. For instance, URLs and mentions are included in the message field, but are then enumerated more particular outside of it - they are considered metadata by SynX. This makes sense because in both SynX and SynTel, the LLM is not given leeway to insert references to URLs, nor is it given enough information about all possible actors for it to determine appropriate mentions. Indeed, recall the dangers of this approach from Chang, et al. in 2024, where requiring the LLM to know about all other nodes raised scalability issues.[24] Instead, both URLs and mentions are handled by the simulation and considered metadata rather than parts of the message. SynX therefore makes changes to the metadata of the original message based on the BEND maneuvers. Table 6.2 illustrates which BEND maneuvers are derivative (associated

with retweets, replies, quotes), which maneuvers are associated with the presence of mentions and which are associated with the presence of URLs.

Table 6.2: Metadata by BEND Category

	<b>Derivative</b>	<b>Mentions</b>	<b>URLs</b>
Bridge		✓	
Build		✓	
Boost		✓	
Back	✓	✓	
Engage	✓		✓
Explain			✓
Excite			
Enhance			✓
Negate		✓	
Neutralize	✓	✓	
Narrow			
Neglect		No Mentions	
Dismiss			
Distort			
Dismay			
Distract			

Once the metadata adjustments are made, a prompt is constructed in preparation to request a message from the LLM. Prompt construction has two major components - the system prompt and the user prompt. The system prompt is used to outline the role given to the LLM and provide background information and context. It has four major parts:

- *An introduction:* You will be participating in a role playing game to help users identify misinformation, disinformation, and manipulation on social media. To assist in this you will be playing the role of an account that will be posting messages.
- *Formatting instructions:* 'Provide your response as a JSON object in the following example format: { "topic": "dogs", "hashtags" : [ "yaydogs", "dogscool" ], "full\_text" : "Dogs are great, #yaydogs #dogscool", "refuse\_to\_answer" : 0 } The "refuse\_to\_answer" field is where you should return a 1 if you do not feel comfortable generating a tweet about the subject. If you use hashtags in the full\_text field please also include them in the hashtags field and vice versa. Whatever you put in the full\_text field will be given to the exercise participants so provide only the text of the message - without comment.
- *BEND Definitions:* As you craft the message/tweet you are trying to accomplish something - what you are trying to accomplish is defined by the BEND Framework. BEND is a framework for describing social-cyber maneuvers. BEND includes 16 different maneuvers. These 16 maneuvers have the following definitions: The BUILD maneuver primarily creates a community. The BACK maneuver primarily increases the importance or effectiveness of a leader... *other maneuvers...*

- *Identity*: You are a Twitter user who is trying to make a post that will be engaging and interesting to your followers. You have a unique style and voice that you want to maintain in your posts. Here are your personal details: {Name}, {Title}, {Age}, {Race}, {Gender}, {Nationality}, {Biography}.

The user prompt is where the LLM is given details about this specific message:

- *Narrative*: You will be posting on the following narrative: {narrative description}
- *Last three messages by this account*: The last three messages you posted looked like this - say something different than these: *Last three messages here...*
- *Last three messages on this narrative*: The last three messages others posted on this narrative looked like this - say something different than these: *Last three messages from narrative here...*
- *Suggested Hashtags*: *List of suggested hashtags...*
- *BEND Maneuvers*: "The message you send will include some BEND maneuvers. In this case: *List of BEND maneuvers...*

The system and user prompts are sent to the LLM and a message is returned. Ideally, the returned message is in the correct JSON format and the LLM has not refused to create the message due to subject matter. No effort was made to jailbreak any LLM in the work done for this thesis - if the OpenAI commercial LLM was unwilling to create a tweet, then SynX defaults to a locally run LLM. A more comprehensive breakdown of SynX prompt construction can be found in Appendix D

The returned text is added to the tweet and there is now a complete post. However, there is no guarantee that the message returned by the LLM contains the intended BEND messages. Because SynX is creating synthetic data for training on BEND maneuvers, ideally it should check the synthetic data for those BEND maneuvers using the same tools that the training audience would have available to them. Thus, SynX uses a combination of NetMapper and ORA-Pro to check for BEND maneuvers in each message. For scalability purposes, SynX can evaluate all posts in an hour together or wait and conduct the evaluation on a full day at a time.

## BEND Check Process

- Messages are cleaned for processing by NetMapper
- NetMapper processes the posts and returns a .tsv of cues per message
- The posts are converted into DyNetML (XML) format for processing by ORA
- The cues .tsv is parsed and the cues injected into the DyNetML
- The BEND calculations are done using templates from ORA-Pro's batch mode
- Each message now has vectors for intended BEND and detected BEND

This process is computationally expensive, therefore, SynX can simultaneously ask the LLM for multiple variations of a message - evaluating and then keeping only the message with the highest score. The scoring algorithm requires that the detected BEND maneuvers in a message

at least encompass the intended maneuvers. Then it gives higher scores to those messages that have the fewest detected BEND maneuvers that are not in the intended BEND maneuvers set.

Let:

- $\mathbf{i} = [i_1, i_2, \dots, i_{16}]$  be the intended BEND maneuver vector
- $\mathbf{d} = [d_1, d_2, \dots, d_{16}]$  be the detected BEND maneuver vector
- $i_k, d_k \in \{0, 1\}$  for  $k = 1, 2, \dots, 16$

Define the score  $S(\mathbf{i}, \mathbf{d})$  as:

$$S(\mathbf{i}, \mathbf{d}) = \begin{cases} 1 - \frac{\sum_{k=1}^{16} [d_k = 1 \wedge i_k = 0]}{16}, & \text{if } \forall k, i_k \leq d_k \\ 0, & \text{otherwise} \end{cases}$$

That is:

- If all intended maneuvers are present in the detected set ( $i_k \leq d_k$  for all  $k$ ),
- then subtract the proportion of "extra" detected maneuvers from 1,
- else, assign a score of 0.

If no suitable messages are found within the batch (all have scores of 0), then SynX will ask for an entirely new batch of messages up to three times. Ultimately, only the best message is kept and added to the corpus.

## E) Find Another Message

If the BEND maneuvers suggest a derivative message - a retweet, reply, or quote - then SynX will need to find an appropriate message to be derivative of. The first step in this process is to determine a subset of messages available for derivative use. This step is done for each hour all at once - reducing computational requirements.

Messages less than or equal to 1 day old have a 100 percent probability of remaining available for derivative use. For older messages, the probability decays exponentially based on age, with the base rate applied to the power of the number of days old the message is - 1.

$$P(\text{keep}) = \begin{cases} 1 & \text{if } \Delta t \leq 1 \\ \beta^{(\Delta t - 1)} & \text{if } \Delta t > 1 \end{cases} \quad (6.1)$$

With a default base keep rate of 0.5, this creates a probability that halves with each additional day of age beyond the first day. This ensures that accounts are generally making derivative messages based upon more recent tweets, while still allowing for significant reach back. Derivative tweets can be eligible for derivative use, but if selected the original tweet and not the derivative is used instead. This means that the derivative use of a tweet essentially refreshes its age out timer.

Recall, that the derivative tweet already has a narrative, therefore, the subset of tweets available for derivative use this hour is further narrowed (for this tweet) by eliminating all tweets that do not share this narrative.

From this shared-narrative set of eligible tweets, SynX will make a selection based on a hybrid of leader-based selection and preferential attachment. The intent is that the network should reflect the scale-free structure provided by preferential attachment; however, the beneficiaries of that scale-freeness need to be the enumerated leaders provided by the group synthetic template from AESOP.

Let  $T$  be the set of candidate tweets, where each tweet  $t \in T$  has an associated retweet/reply/quote count  $R(t) \geq 0$ , and a leader indicator  $L(t) \in \{0, 1\}$ , where  $L(t) = 1$  if the tweet was posted by a leader of the group.

Define  $T_L \subseteq T$  as the subset of tweets authored by leaders:

$$T_L = \{t \in T \mid L(t) = 1\}$$

We define the probability  $P(t)$  of selecting tweet  $t$  as follows:

$$P(t) = \begin{cases} \frac{1}{|T_L|} & \text{with probability } \frac{1}{3}, \quad \text{if } t \in T_L \\ \frac{R(t)}{\sum_{s \in T} R(s)} & \text{with probability } \frac{2}{3}, \quad \text{for all } t \in T \end{cases}$$

To sample a tweet  $t^*$ , first choose a mode of selection:

- With probability  $\frac{1}{3}$ , sample uniformly from  $T_L$ .
- With probability  $\frac{2}{3}$ , sample from  $T$  using retweet counts as weights.

Once a tweet is chosen, metadata is adjusted as appropriate (see D above). If the derived tweet is a retweet then most of the intended BEND maneuvers for the derivative tweet are overwritten with the intended BEND maneuvers from the original tweet - the exceptions being back, engage, and neutralize - the determining BEND factors for derivatives. However, if the derived tweet is a reply or quote, then it keeps its intended BEND maneuvers and moves to a modified version of step D above - where the LLM is asked to comment on the tweet derived from.

## F) Produce Output

Finally, the tweet, retweet, reply, or quote is complete and is added to the full set of output tweets for the hour. These tweets will be added to the list of available for derivative use tweets in the next iteration. The entire process is repeated for each hour of the exercise.

## 6.4 SynTel

SynTel executes in a similar manner to SynX - with roughly analogous steps A-F - see Fig. 6.5. However, the Telegram platform has structural differences that introduce some changes. In SynTel, channels and user accounts are handled simultaneously, with user accounts adding original posts to their own channel, cross-posting from another channel, or posting to another channel rather than choosing an option from tweet, retweet, reply, or quote. This causes SynTel Telegram

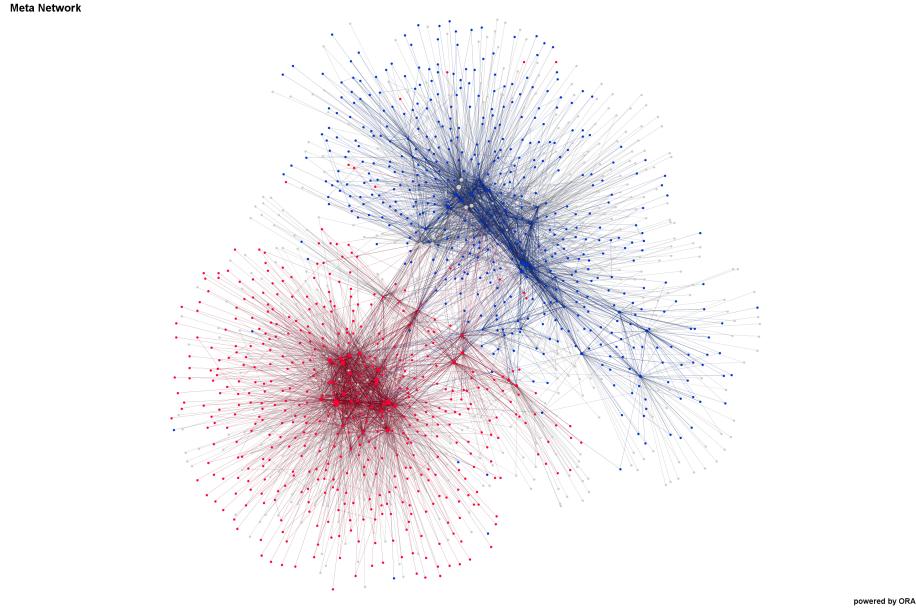


Figure 6.4: An output network from SynX with pro-US stance detection run through ORA-Pro. Blue nodes are pro-US actors and red nodes are anti-US actors with gray nodes being neutral. Nodes are sized by total degree centrality.

networks to be more condensed. Additionally, while mentions and hashtags are supported for Telegram creation in SynTel, Telegram users traditionally use these features less, and therefore shared URLs and cross-posting become more dominant in network structures. SynTel accounts for this by increasing the use of URLs in Telegram messages while reducing the likelihood of mentions and hashtags.

## X/Twitter and Telegram Bots

No social network would be complete without bots. However, there is an unusual problem with synthetic datasets in bots - every actor in this fabricated information environment except the training audience is an automated persona - it is all bots. The challenge is that for the training audience, some of these automated personas need to be detectable as "bots" and others need to remain undetected as "humans". Planners can enumerate bots in AESOP for both SynTel and SynX to simulate and both have agent rule sets for amplifier, news, bridging, and repeater bots. Amplifier bots boost content through retweets/posts and the SynTel/X rule-set ensures that amplifier bots have an abnormally high retweet/post to tweet/post ratio. Additionally, they target only content with a specified narrative and ignore the recency bias imposed by SynTel/X on normal actors. Repeater bots are similar to amplifier bots but repeat the same message within their in-group continuously. Both of these types of bots operate similarly to those seen by Ng and Carley in 2023[57]. News bots are news aggregators and have a rule-set that forces them to retweet/post content from a target set of news agencies. Bridging bots use mentions and retweets/posts to attempt to connect two specified narratives. These types of bots operate similarly to those observed

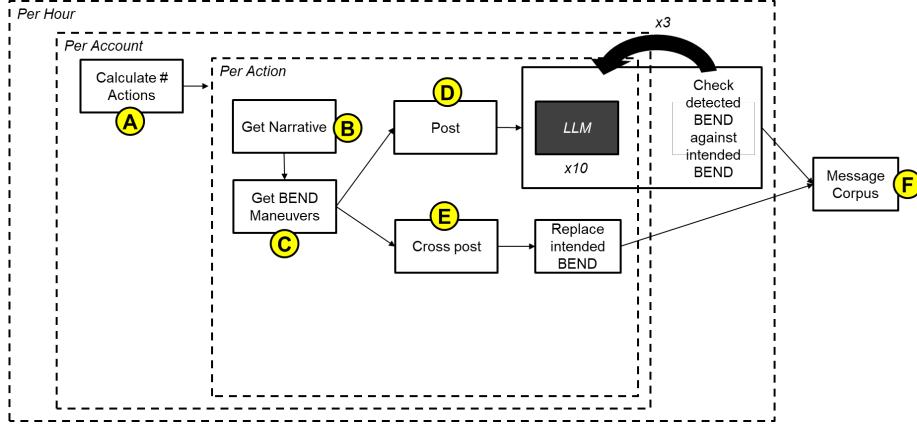


Figure 6.5: Telegram generator message creation flow. Dark gray parallelograms are synthetic templates provided by AESOP and referenced by the generator.

by Jacobs et al. in 2023 [42]

## 6.5 Validation

MOMUS, a Netanomics product, is custom made to accept the scenarios from AESOP and the output from SynTel/X and ensure that the data conforms to the scenario and network and narrative structure fall within the norms for a platform. Unfortunately, MOMUS is still in development and not fully capable of providing independent verification of the synthetic data at the time this research was being conducted. Therefore, while initial MOMUS results are included, a different method was required to validate the synthetic data.

The purpose of the data is to enable BEND maneuver training on social media. Since AESOP ensures that there are BEND maneuvers present within the synthetic templates, a straightforward method for validating the corpus generated off of those synthetic templates would be to check that the intended BEND maneuvers are present in the output. If the training audience is able to find the intended maneuvers then the data has accomplished its training purpose.

Therefore, this validation takes two separate forms:

### 1) Validate Overall Reasonableness of Data - OR - Can analyst tools meaningfully run on the data?

In addition to facial validity supported by the design decision alignments found in the stylized facts in Table 6.3, additional validation comes from directly comparing network metrics between the outputs of SynX - run on an AESOP generated scenario - and a real-world comparison dataset. The real-world comparison dataset is approximately 2500 tweets pulled from the X/Twitter API during the Balikatan 22 exercise in April 2022. Balikatan is a bilateral military exercise between the US and the Philippines.[52] The input scenario from AESOP is meant to be similar in topic and content.

I evaluated SynX outputs on five different network metrics against Balikatan 22. Three were

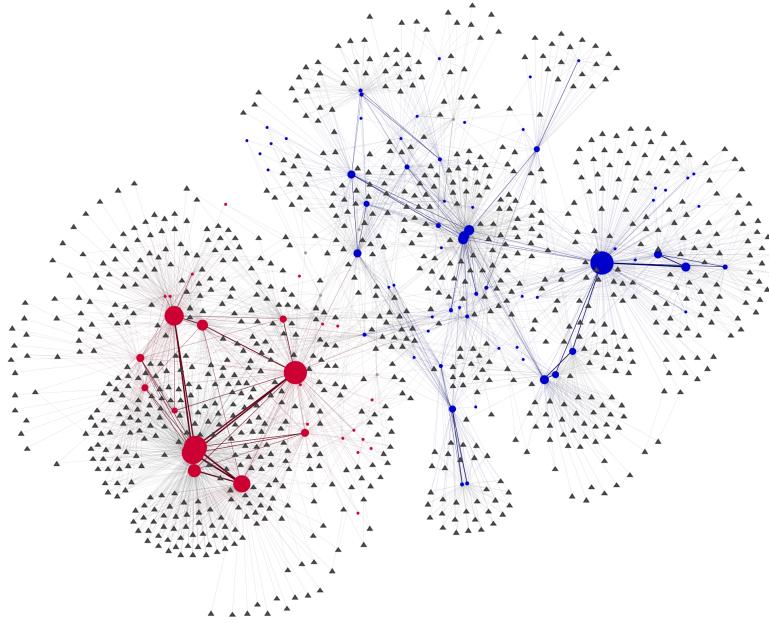


Figure 6.6: An output network from SynTel with pro-US stance detection run through ORA-Pro. Blue nodes are pro-US actors and channels. Red nodes are anti-US actors and channels. Triangle shaped nodes are messages. Nodes are sized by total degree centrality.

suggested by Chang et al. in 2024 during their development of an LLM-based synthetic network simulation.[24] I added Degree Distribution of the Tweet x Tweet (Retweet) network because ORA-Pro's primary method for stance propagation depends upon this network being well-connected. It is not necessary that SynX be precisely equivalent to the Balikatan 22 scenario but that the differences should be plausible. We are more concerned with matching shape and form than exact values. The four metrics are:

- Average Shortest Path of Agent x Agent (All Communication)
- Proportion of nodes in the Largest Connected Component of Agent x Agent (All Communication)
- Modularity of Agent x Agent (All Communication)
- Degree Distribution of Tweet x Tweet (Retweet)

From the results in Table 6.4 we can see that SynX falls within acceptable ranges (within 10%) for two of the three Agent x Agent network metrics when compared with the Balikatan 22 real dataset. SynX produces data sets that have agents that are too modular compared to the example set. This is likely due to the constraints placed upon SynX by adhering to the AESOP scenario which has rigidly defined narrative groups. Regardless, more work will need to be done to improve AESOP and SynX in this area.

Finally, when looking at the Tweet x Tweet (Retweet) network results we can see that Synx faithfully recreates a scale-free network and closely mimics the Balikatan network - see Table

Table 6.3: Stylized Facts

Summary	Effect	Source
Attention spans limit how many users are affected	Reach is determined by leadership	Kang and Lerman, 2013. [49] Lu et al., 2014. [54]
Commonality of reposts	High level of derivative messages (50+%)	Beskow and Carley's "Agent Based Simulation of Bot Disinformation Maneuvers in Twitter" from 2019.[11]
Which messages are being reposted	Power law distribution	Lu et al., 2014. [54]
Real world likelihood of BEND maneuvers	variable-order Markov model with set-based context and non-replacement sampling based on tweets covering the Balikatan 22 exercise in APR 2022	Lepird, 2024.[52]

Table 6.4: Network Metrics Comparison for Agent x Agent Network

Metric	Real	Sim Mean	Sim Std	Abs. Diff.	Perc. Diff. (%)
Normalized Avg. Shortest Path	0.4158	0.4429	0.0294	-0.0272	-6.54
LCC Proportion	0.8738	0.9583	0.0123	-0.0845	-9.67
Modularity	0.5006	0.9174	0.003	-0.4169	-83.28

Table 6.5: Degree Distribution Statistics for Tweet x Tweet Network

Network	Node Count	Mean Degree	Median Degree	Max Degree	Std Dev
Real Data	2230	0.92	0.0	154	6.64
Simulation Avg	10457	0.88	0.0	148	4.74

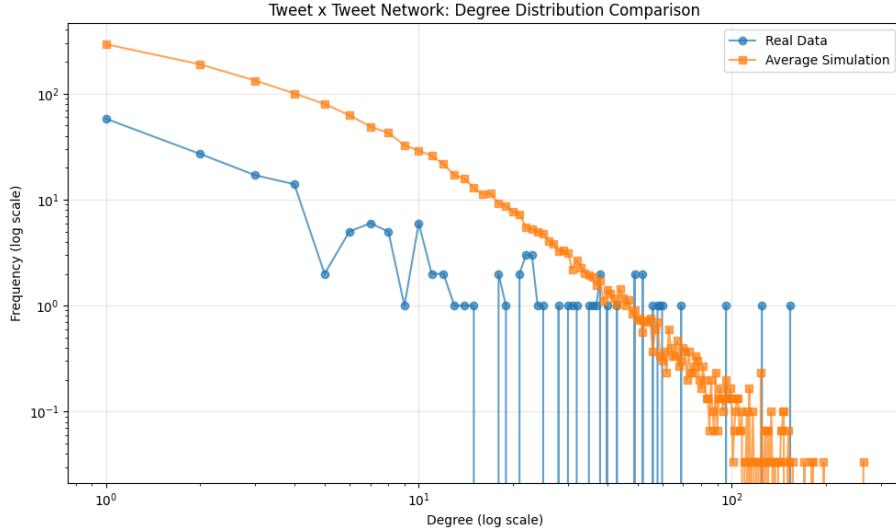


Figure 6.7: Both the real data and the SynX display appropriate power-law properties for the Tweet x Tweet (Retweet) network

6.5 and Figs. 6.7 and 6.8.

## 2) Validate Execution of Intention - OR - Is what I intended to put in the data actually found in the data?

Critical to developing a scenario and accompanying training data for a training audience is understanding the learning objective of the training. In this case, as outlined in Chapter 4, the training focuses on developing social media analysis skills for the detection and application of BEND maneuvers in the information environment. Therefore, the data should include BEND maneuvers for the training audience to find and analyze. I set out to maximize the matching between the BEND maneuvers outlined by the AESOP provided scenario and the BEND maneuvers detected in the dataset output by SynX. To do this, I conducted a virtual experiment to find the optimal method for prompt insertion - validating BEND maneuver output while doing so.

### 6.5.1 Experiment

For the virtual experiment, I will manipulate the presence of the BEND definitions in the prompt and whether SynX attempts to query the LLM for different text based on incorrect BEND assessments. I will be evaluating how well the presence of BEND maneuvers and their co-occurrences match a real world dataset. The real world comparison dataset is approximately 2500 tweets pulled from the X/Twitter API during the Balikatan 22 exercise in April 2022. Balikatan is a bilateral military exercise between the US and the Philippines.[52] The input scenario from AESOP is meant to be similar in topic and content.

In this experiment, I looked at how much benefit is gained by including the comprehensive BEND definitions and the NetMapper cues that ORA-Pro will be looking for when evaluating

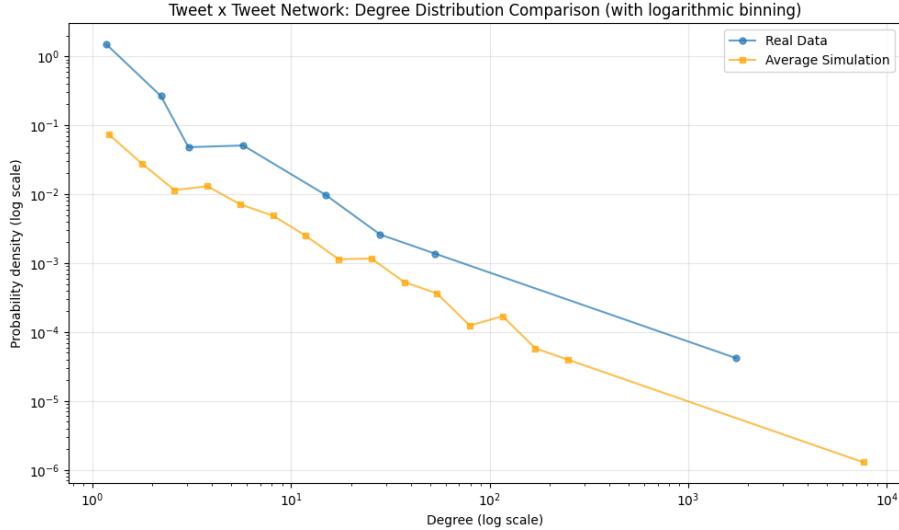


Figure 6.8: Binning is not always a good idea, but with the discontinuous data from the real data this chart helps depict the form convergence between the real and synthetic data sets

BEND in the prompt to the LLM. I consider three cases for the prompt - 1) No BEND definitions, 2) BEND definitions fit into the user prompt to the LLM, and 3) BEND definitions in the system prompt to the LLM. The user prompt is generally understood to be the prompt that asks the question, while the system prompt is the prompt that lets the LLM general context and purpose.

I also investigated iterating with the LLM multiple times to get better BEND results by letting the LLM know what BEND maneuvers were not detected in its initial responses and re-iterating the associated BEND cues with those maneuvers.

I will run SynX on one hour of the AESOP scenario (1200-1300 on Day 6 of the exercise) producing between 600-800 messages.

### 6.5.2 Results

For the results, we can analyze both the presence of BEND maneuvers and the mixture of these maneuvers by creating a co-occurrence matrix. Fig 6.9 gives an example of the heatmap from a single run of the experiment. Notice that it is a comparison between the intended BEND maneuvers - those requested from the LLM - and the detected BEND in the results.

It is obvious from this that SynX is not perfect - there should be ones in the diagonal such that the maneuvers we ask for the LLM returns. However, it is also important to note that where a BEND maneuver is detected as something else - everywhere else but the diagonal - this is not necessarily bad. There is some level of co-detection and co-occurrence naturally even in real data, i.e. rarely do maneuvers occur in isolation independent of one another. In order to account for this we need to use the same correlation matrix derived from real data - specifically the Balikatan 22 X/Twitter dataset. We take the mean co-occurrence for each day and construct a standard deviation for each combination.

This allows us to then compare a single run of results and get a distance from the mean in

Table 6.6: 3x2 Virtual Experiment

Independent Variables	# Test Cases	Values Used
BEND Definitions	3	None/User_Prompt/System_Prompt
Regen Messages for BEND	2	0x/3x
Control Variables	# Test Cases	Values Used
Time Periods	1	1 hour
Messages per Hour	1	600-800
Target BEND	1	Balikatan 22
AESOP Input	1	Scenario
Dependent Variables		Values Expected
BEND Distribution		0-1 z-scores
3x2 Replications per cell		8 cells 30 180 total runs (108,000+ messages)

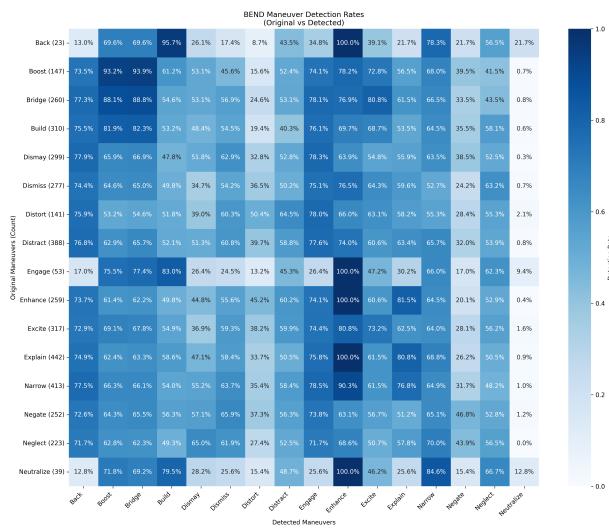


Figure 6.9: Example co-occurrence heatmap (that shows intended BEND maneuvers mapped to detected BEND maneuvers).

Table 6.7: Overall BEND Maneuver Analysis Results

Experiment	Mean Z-Score	Standard Error	CI Lower	CI Upper	p-value
No Definitions	-0.3366	0.0252	-0.3881	-0.2851	0.000000
Definitions_user	-0.4053	0.0270	-0.4605	-0.3501	0.000000
Definitions_system	-0.1053	0.0263	-0.1592	-0.0515	0.000396
Definitions_repull	-0.1257	0.0218	-0.1703	-0.0812	0.000003

Significance determined at  $p < 0.05$  level.

Z-scores represent deviation from expected co-occurrence patterns in real data.

Table 6.8: BEND Maneuver Analysis - Intended/Requested (Row Means)

Maneuver	No Definitions	Definitions_user	Definitions_system	Definitions_repull
back	<b>0.35<sup>+</sup></b>	<b>-0.42<sup>-</sup></b>	<b>-0.26<sup>-</sup></b>	<b>-0.22<sup>-</sup></b>
boost	<b>-0.67<sup>-</sup></b>	<b>-0.89<sup>-</sup></b>	<b>-0.69<sup>-</sup></b>	<b>-0.66<sup>-</sup></b>
bridge	<b>-0.44<sup>-</sup></b>	<b>-0.38<sup>-</sup></b>	-0.10	-0.05
build	-0.24	-0.07	0.17	0.21
dismay	<b>-0.87<sup>-</sup></b>	<b>-0.80<sup>-</sup></b>	<b>-0.35<sup>-</sup></b>	<b>-0.50<sup>-</sup></b>
dismiss	-0.21	-0.11	<b>0.17<sup>+</sup></b>	0.11
distort	<b>-0.71<sup>-</sup></b>	<b>-0.65<sup>-</sup></b>	-0.29	<b>-0.31<sup>-</sup></b>
distract	<b>-0.73<sup>-</sup></b>	<b>-0.64<sup>-</sup></b>	<b>-0.31<sup>-</sup></b>	<b>-0.34<sup>-</sup></b>
engage	-0.12	<b>-0.47<sup>-</sup></b>	<b>-0.28<sup>-</sup></b>	<b>-0.24<sup>-</sup></b>
enhance	0.11	0.17	<b>0.45<sup>+</sup></b>	<b>0.39<sup>+</sup></b>
excite	<b>-0.64<sup>-</sup></b>	<b>-0.71<sup>-</sup></b>	<b>-0.39<sup>-</sup></b>	<b>-0.44<sup>-</sup></b>
explain	<b>-0.46<sup>-</sup></b>	<b>-0.46<sup>-</sup></b>	0.04	-0.06
narrow	<b>-0.41<sup>-</sup></b>	<b>-0.30<sup>-</sup></b>	0.13	0.13
negate	<b>-1.00<sup>-</sup></b>	<b>-1.08<sup>-</sup></b>	<b>-0.70<sup>-</sup></b>	<b>-0.76<sup>-</sup></b>
neglect	<b>0.98<sup>+</sup></b>	<b>0.95<sup>+</sup></b>	<b>1.22<sup>+</sup></b>	<b>1.18<sup>+</sup></b>
neutralize	<b>-0.33<sup>-</sup></b>	<b>-0.63<sup>-</sup></b>	<b>-0.50<sup>-</sup></b>	<b>-0.44<sup>-</sup></b>

terms of standard deviation. We can also put the runs from a single experimental set-up together to see how well any given set-up does, as in Fig. 6.11.

We can also compare the average z-score from the runs within an experimental setup. Fig. 6.12 shows an example of these results. Finally, we can compare the overall distance from the mean across the four main configurations as in Fig. 6.13. From Fig. 6.13 we can determine that overall, SynX under-represents BEND maneuvers within the synthetic messages. Also, there was no significant difference between including the BEND definitions in the user prompt versus leaving them out entirely. This might be due to the relative proximity of extraneous definitions - i.e. those BEND maneuvers not being requested for this particular message - with the BEND maneuver requests specific to the message. However, placing the definitions in the system prompt - increasing the distance between overall BEND definitions and the specific ones being requested - significantly increases the accuracy of the LLM returned responses. However, the methodology we take for trying to re-request the LLM for a better BEND response provided no significant improvement.

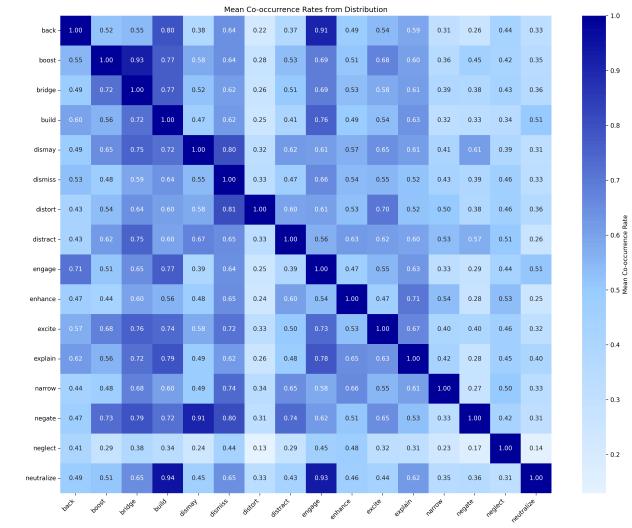


Figure 6.10: Co-occurrence heatmap of the Balikatan 22 data - note that there is no concept of intended vs detected BEND maneuvers in real data

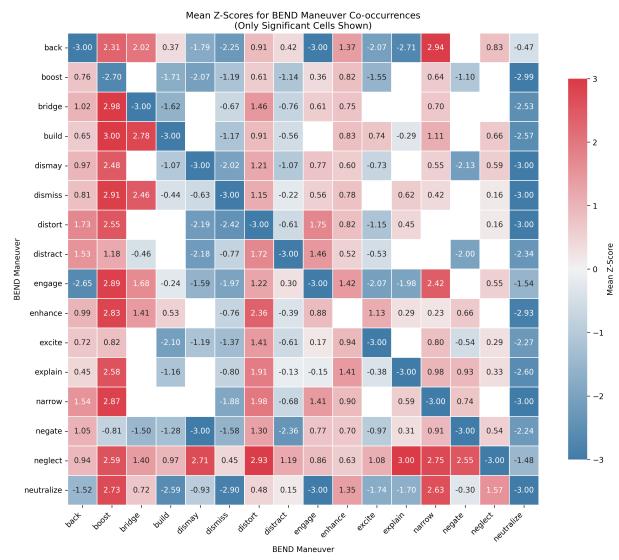


Figure 6.11: Co-occurrence matrix across all runs that had the BEND definitions in the system prompt but did not repoll the LLM

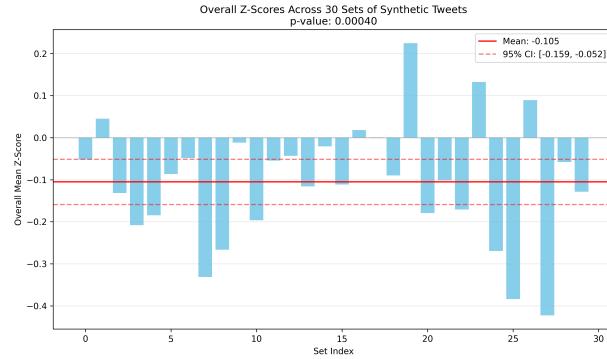


Figure 6.12: A comparison of the overall distance from the mean for all runs that had the BEND definitions in the system prompt but did not repoll the LLM

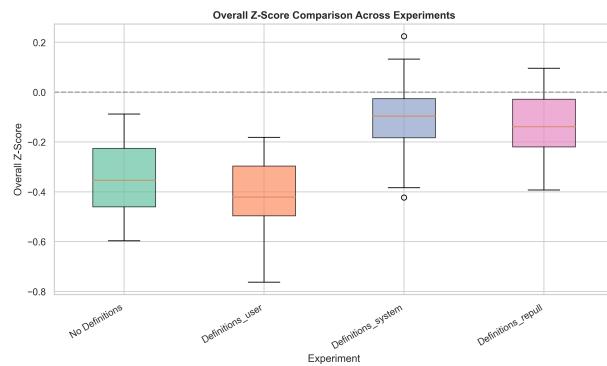


Figure 6.13: A comparison of the overall distance from the mean across all BEND maneuvers from each of the four major experiment setups

Table 6.9: BEND Maneuver Analysis - Detected (Column Means)

Maneuver	No Definitions	Definitions_user	Definitions_system	Definitions_repull
back	<b>0.35</b> +	<b>0.36</b> +	<b>0.37</b> +	<b>0.37</b> +
boost	<b>-1.24</b> -	<b>1.20</b> +	<b>1.95</b> +	<b>1.85</b> +
bridge	<b>-1.27</b> -	<b>-0.55</b> -	<b>0.51</b> +	<b>0.43</b> +
build	<b>-1.83</b> -	<b>-1.34</b> -	<b>-0.85</b> -	<b>-0.75</b> -
dismay	<b>-0.42</b> -	<b>-1.47</b> -	<b>-1.02</b> -	<b>-1.13</b> -
dismiss	<b>-2.35</b> -	<b>-1.96</b> -	<b>-1.52</b> -	<b>-1.61</b> -
distort	<b>1.14</b> +	<b>1.21</b> +	<b>1.16</b> +	<b>1.04</b> +
distract	<b>0.34</b> +	<b>-0.76</b> -	<b>-0.59</b> -	<b>-0.56</b> -
engage	-0.02	-0.01	0.02	0.02
enhance	<b>0.85</b> +	<b>0.85</b> +	<b>0.87</b> +	<b>0.85</b> +
excite	<b>-0.83</b> -	<b>-1.43</b> -	<b>-0.65</b> -	-0.30
explain	<b>-0.41</b> -	<b>-0.61</b> -	<b>-0.29</b> -	<b>-0.35</b> -
narrow	0.18	<b>0.51</b> +	<b>0.89</b> +	<b>0.83</b> +
negate	<b>1.38</b> +	<b>-0.46</b> -	-0.26	<b>-0.52</b> -
neglect	<b>1.16</b> +	<b>0.43</b> +	<b>0.17</b> +	<b>0.25</b> +
neutralize	<b>-2.42</b> -	<b>-2.43</b> -	<b>-2.43</b> -	<b>-2.43</b> -

## 6.6 Future Work and Limitations

There are significant limitations to the current model that can be corrected in future work. First, the agents in the current model have opinions matrices that map their stance on each topic in the scenario; however, there is currently no population modeling implemented to feed these matrices nor is there information diffusion implemented such that these stances change over time. Thus, while the actors respond in accordance with their opinions, the simulation is currently a perfect model of arguing on Internet forums - no one ever changes their mind.

Furthermore, more validation should be done. If the goal is simply to train on BEND maneuver detection and response, then the current validation network metrics and BEND evaluations show that SynX is sufficient. However, training and instruction on other aspects of social networks might require additional, unevaluated, network properties. This includes more research into matching the Agent x Agent modularity against real world datasets. Netanomics does have a fully featured evaluation tool for synthetic social media data sets, MOMUS. MOMUS will verify the synthetic dataset's adherence to the AESOP scenario, the semantic and syntactic content of the messages, and the overall network structure against real world data. This will provide a more thorough evaluation of the dataset. These same evaluations should be mapped onto SynTel. These results should confirm what is currently only assumed - prompting for a post from an LLM told that it is to emulate a Telegram user renders similar results to prompting for a post from an LLM that is told it is to emulate an X/Twitter user.

Additional work should be done to improve the re-request methodology from the LLM. We provided the same prompt to the LLM again but added the original returned results as well as an explanation of what was present and what was missing. Other methodologies should be explored, including asking the LLM itself to improve the message.

Finally, work is also required within the BEND evaluations. While the AESOP scenario comes with images and news sites, the current ORA-Pro and NetMapper BEND evaluation is done only on the meta-data and the text. Future detection should include an evaluation of the included images or referenced URLs.

## 6.7 Conclusions

I created an agent-based simulation, SynX, to model synthetic social media based on input scenario templates from Netanomics' AESOP tool. Validation testing shows that SynX is capable of creating highly dynamic datasets that present BEND maneuvers similar to those of real-world datasets. Additionally, I experimented with different techniques for interaction with an LLM and found that the inclusion of BEND definitions in system prompts was most effective. The validation and results show that SynX provides an effective way to create tailored social media datasets for analyst training on BEND maneuvers.

# Chapter 7

## BEND: Effects-based detection

Social media is organized into Topic-Oriented-Communities . Groups of actors who talk to each other about the same topic. They can be analyzed using network analysis. They can be changed at both the narrative and community level by information maneuvers.

### 7.1 Research Questions

ORA and NetMapper combine to provide a BEND report that automatically detects BEND maneuvers. This methodology is based on work by Blane, who laid out a framework for analysis that uses a complex method for weighting CUES to identify maneuvers.[13] CUES here refers to the linguistic cues extracted from the message text through the NetMapper software. These linguistic cues are what ORA-Pro uses to identify BEND maneuvers per message text.

In this chapter, I propose to move beyond identifying BEND maneuvers within specific messages based on derived intent. Instead of taking a source and a message and extrapolating an intended effect, I want to identify effects experienced by a target in order to determine the BEND maneuver experienced. To use an analogy from military targeting, instead of looking at the aircraft and the weapons load-out and determining that this will be a mission to destroy a tank, I want to look at the target, in this case the tank, and determine exactly what effects it experienced. This approach is captured in Figure 7.1.

This will require bounded, over-time comparisons of groups in order to detect changes in important network and narrative metrics. I will also need to account for more than one effect occurring at a time. Additionally, target identification - especially group target identification is an outstanding issue. Even more importantly, this thesis will require network and narrative metrics tied to the effects of the BEND maneuvers.

The key research questions for this chapter are:

- How can we detect the presence of BEND maneuvers through their effects?
- Can we match these maneuvers with narrative campaigns?

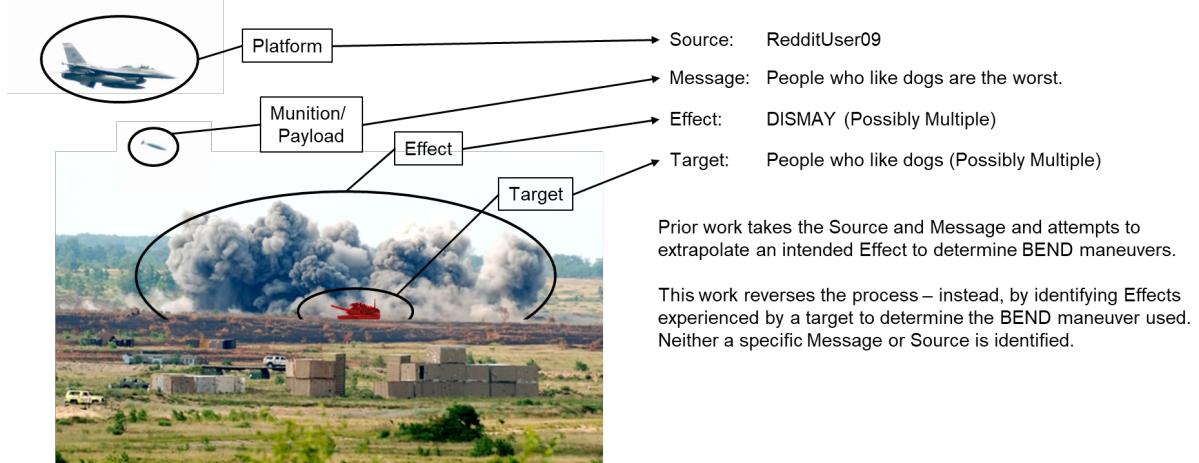


Figure 7.1: An analogy for how intent-based BEND and effects-based BEND interact

## 7.2 BEND-Effects Definitions

I have developed a set of network and narrative metrics to define BEND effects for X/Twitter. These effects Generally these metrics involve changes over time above the baseline corpus - requiring a computation of the metric against both the corpus and the target/target group. Additionally, many require heterogeneous graphs - involving actors, topics, and stance. This approach requires strict definitions for the entities involved and metrics derived.

First, I define two types of topic-oriented groups (TOGs). Because the BEND framework involves maneuvers in both the network and narrative spaces, these groups need to be cohesive in both. This is what makes them topic-oriented groups. In general, grouping of agents is done in network space - here network refers to typical connections gleaned from social media corpora - retweets, quotes, replies, mentions, etc. Therefore, they are groups but I also need them to be narratively connected - they need to not just be groups but *topic-oriented* groups.

Therefore, a topic-oriented group (TOG) is a group of agents clustered in both network and narrative space. However, because all of the BEND-effects are over-time changes there is a distinction between two related topic-oriented groups. A topic-oriented cluster (TOG cluster) is a TOG found within a single time-frame. Matching multiple topic-oriented clusters through time reveals communities that stretch across the corpus in time - these are called topic-oriented communities (TOG community). Figure 7.2 shows the key I will be using for illustrations in the following sections.

- Topic-Oriented Group (TOG): a group clustered in both the network and narrative space
- TOG cluster: an identified topic-oriented group in a single time frame
- TOG community: linked topic-oriented clusters through time

With these in mind I will now provide the standard BEND definitions derived from Blane[13] and pair them with their corresponding effects-based definition and metrics.

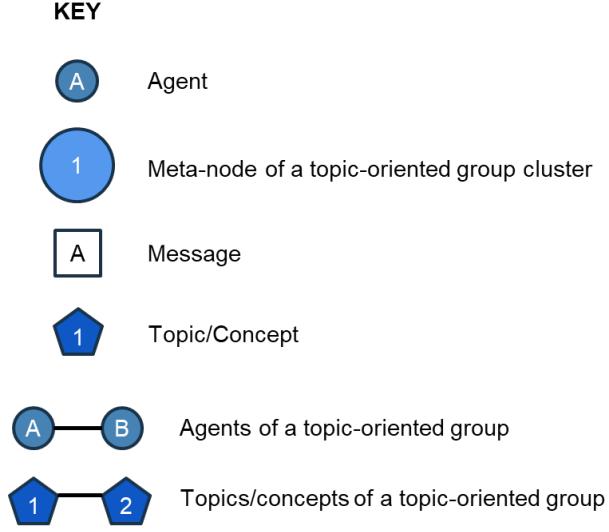


Figure 7.2: BEND-Effects key for illustrations

### Bridge/Neutralize

**Bridge:** Discussion or actions that establish a connection between two or more groups or create the appearance of such a connection [13]

**Neutralize:** Discussion or actions that cause a group to be dismantled or appear to be dismantled [13] *[Note: This was originally titled "Nuke" in Blane's thesis but was changed to prevent confusion and amended to now translate as: Discussions or actions that cause a group to be, or appear to be, no longer of relevance, e.g. because it was dismantled.]*

Bridge maneuvers involve discussion or actions that build a connection between two or more groups or create the appearance of such a connection. To detect this, I will look at the overall connections between between any two clusters and their change over time. If the number of links between the representative meta-nodes of any two clusters increases over time then bridging is occurring. Conversely, if the number of links is decreasing then neutralization is occurring.

To calculate this, I let  $C_i$  and  $C_j$  denote two clusters of nodes identified in time period  $t_1$ . Let these same sets of nodes be examined in time period  $t_2$ .

Define:

- $L_{ij}^{(t_1)}$ : the number of links between nodes in clusters  $C_i$  and  $C_j$  at time  $t_1$
- $L_{ij}^{(t_2)}$ : the number of links between the same clusters  $C_i$  and  $C_j$  at time  $t_2$
- $L_{ij}^{\text{total}} = L_{ij}^{(t_1)} + L_{ij}^{(t_2)}$ : the total number of links between the two clusters across both time periods

We normalize the inter-cluster links in each time period as follows:

$$\hat{L}_{ij}^{(t_1)} = \frac{L_{ij}^{(t_1)}}{L_{ij}^{\text{total}}}, \quad \hat{L}_{ij}^{(t_2)} = \frac{L_{ij}^{(t_2)}}{L_{ij}^{\text{total}}}$$

Then compute the change in normalized inter-cluster connectivity:

$$\Delta_{ij} = \hat{L}_{ij}^{(t_2)} - \hat{L}_{ij}^{(t_1)}$$

Interpretation:

- If  $\Delta_{ij} > 0$ , bridging is **increasing**
- If  $\Delta_{ij} < 0$ , bridging is **decreasing** (neutralizing)
- If  $\Delta_{ij} = 0$ , the connection is **stable**

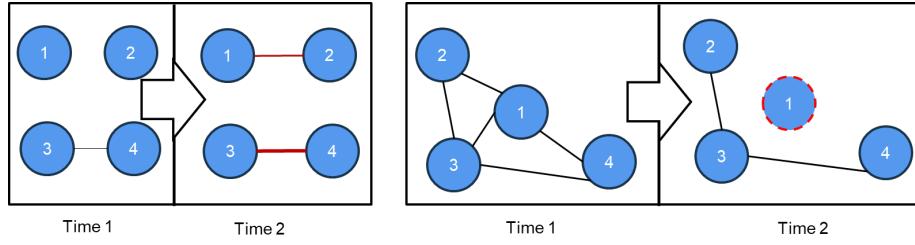


Figure 7.3: Visualization of bridge/neutralize.

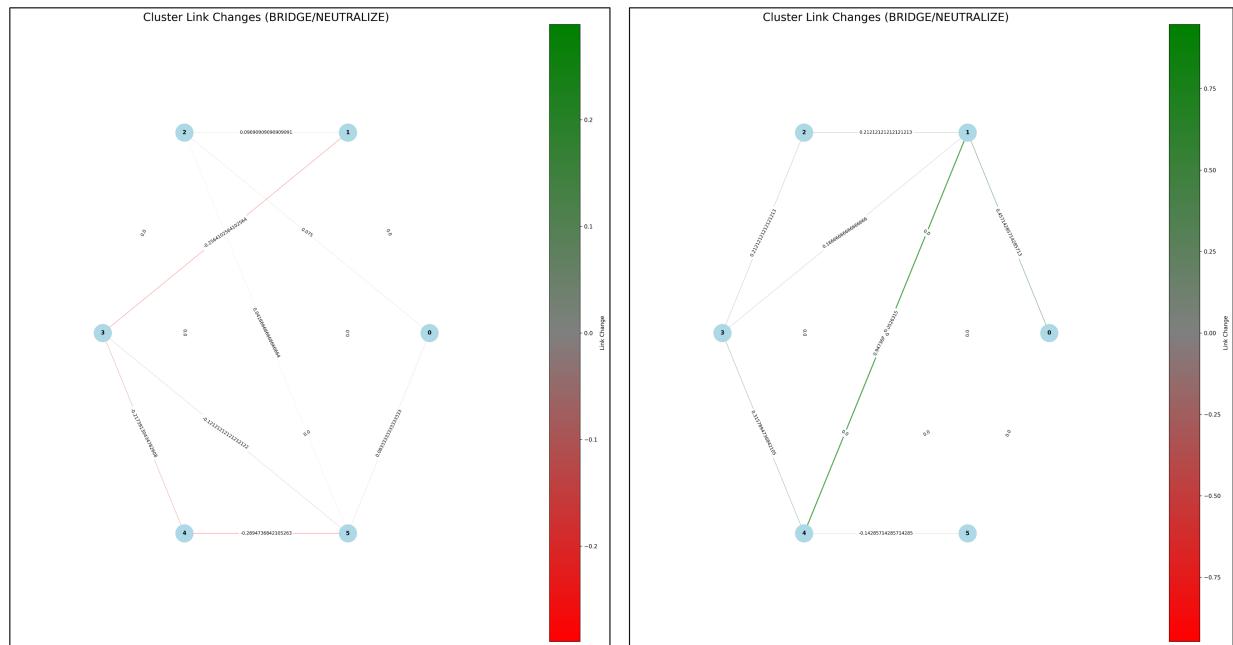


Figure 7.4: Actual results of bridge/neutralize.

## Back/Negate

**Back:** Discussion or actions that increase the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic [13]

**Negate:** Discussion or actions that limit the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic [13] [Note: This was originally "Neutralize" in Blane's thesis but "Neglect" became "Negate" and was switched with "Neutralize" for clarity purposes.]

Backing and negation of actors within a corpus is already an important feature for most network analysis tools and ORA-Pro already provides analytic tools for this purpose. In order to detect if an actor has been the target of a back maneuver, I will be looking for a positive change over time - above the baseline corpus - in the centrality of an actor within the actor interaction network. Inversely, for negate, I will be looking for a negative change in a target actor's centrality over time of a magnitude greater than that of the baseline corpus. ORA-Pro provides Super Spreader, Super Friend, General Influencer, and Eigenvector Centrality metrics that are used in ensemble for this maneuver.

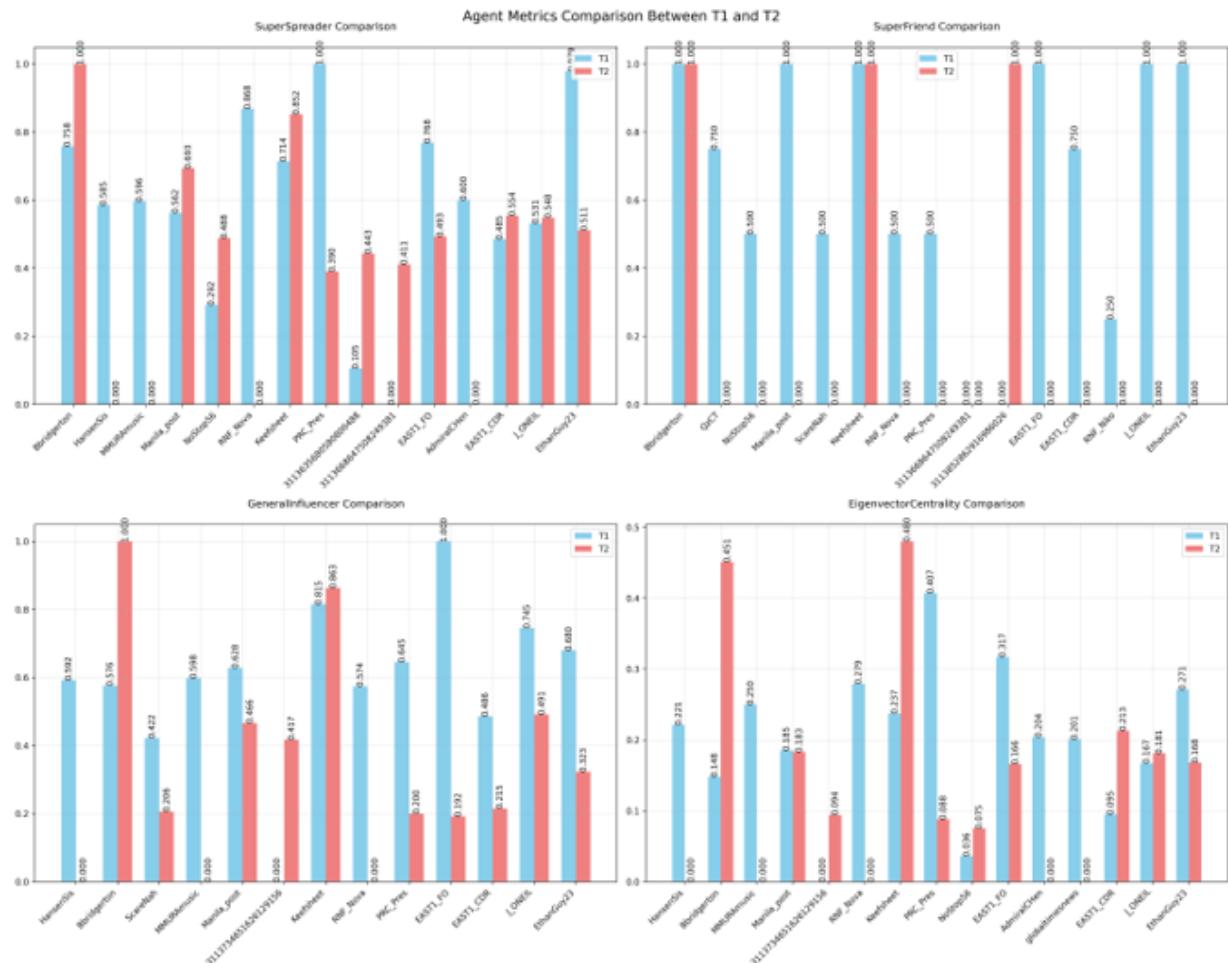


Figure 7.5: Actual results of back/negate.

## Build/Narrow

**Build:** Discussion or actions that create a community or create the appearance of a community [13]

**Narrow:** Discussion or actions that lead a group to fission (or appearance of fission) into two or more distinct groups [13]

Build maneuvers are marked by discussion or actions that create a group, or the appearance of a group, where there was none before. Therefore, I will be looking for either new group emergence - specifically new Leiden-detected topic-oriented clusters under the BEND-effects construction - or multiple smaller groups forming a new, larger community.

Narrow maneuvers exhibit discussion or actions that lead a group to be, or appear to be, more specialized, and possibly to fission, or appear to fission, into two or more distinct groups. Within BEND-effects, I will be looking for a single group splitting into multiple groups or the disappearance of links on the bipartite network from meta-agent group node to topic/stance nodes.

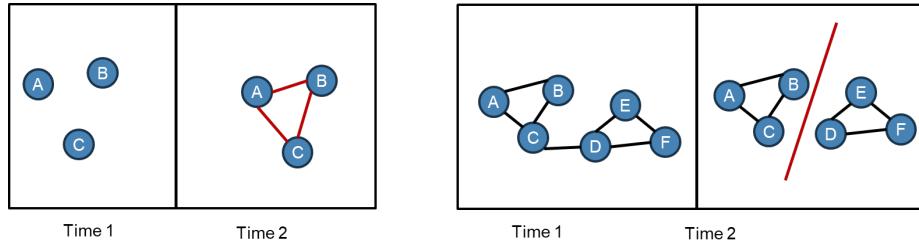


Figure 7.6: Visualization of build/narrow.

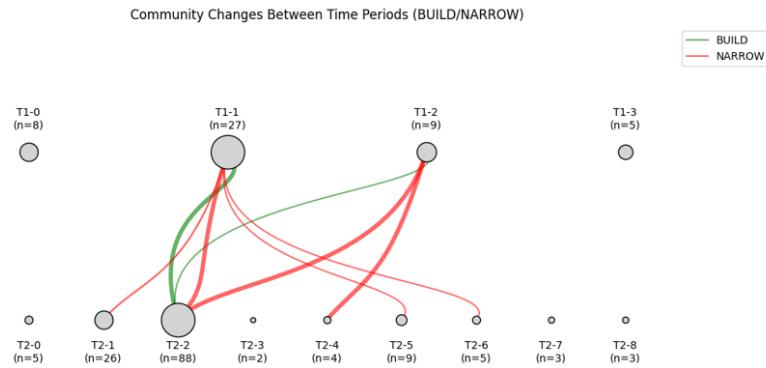


Figure 7.7: Actual results of build/narrow.

## Boost/Neglect

**Boost:** Discussion or actions that increase the size of a group and the connections among group members or the appearance of such [13]

**Neglect:** Discussion or actions that decrease the size of the group, or the connections among the members, or the appearance of these [13]

Boost maneuvers require discussion or actions that increase the size of a group and/or the connections among group members, or the appearance of such. I will look at average path length in the agent x agent network in order to determine if a group is experiencing boosting - a decrease in average path length. A neglect maneuver effect would be an increased average path length.

This approach solves two significant issues. The first is the temptation to use density calculations for boost/neglect. Unfortunately, density is tied closely to network size such that a large increase in nodes in a cluster would reduce density and thereby reduce boosting even though increased node numbers would seem to be a secondary objective of the boosting maneuver. Average path length is more resistant to size changes and therefore a better metric.

The second issue is that both build and boost maneuvers allude to size changes in a topic-oriented group. In differentiating the two, I opted to look carefully at my initially framing - topic-oriented groups via Leiden grouping. Because Leiden grouping allows for singleton clusters, topic-oriented groups can be single nodes. There could not meaningfully be 'new' groups from time to time unless the agents themselves are wholly new. Instead, what we see are groups (often singleton clusters) coming together to form new larger groups - this fits the original definition of 'build' - non-cohesive individuals coming together - despite it really just being a size change/amalgamation of groups under the new topic-oriented groups framing device. This is why I opted to push size solely into the build maneuver and pair its merging with the narrow maneuver's splitting. This leaves the boost maneuver with 'connections among the members' - and therefore recommends average path length as an appropriate metric.

So to calculate boost and neglect, I let  $C_k$  be a community of nodes identified in both time periods  $t_1$  and  $t_2$ , and let  $G_k^{(t)}$  be the subgraph induced by  $C_k$  at time  $t$ , using only internal links (i.e., the agent  $\times$  agent network restricted to nodes in  $C_k$ ).

Define:

- $\ell_k^{(t)}$ : the average shortest path length in  $G_k^{(t)}$

Then compute:

$$\Delta\ell_k = \ell_k^{(t_2)} - \ell_k^{(t_1)}$$

Interpretation:

- If  $\Delta\ell_k < 0$ , then community  $C_k$  is experiencing **boosting** (increased internal cohesion)
- If  $\Delta\ell_k > 0$ , then community  $C_k$  is experiencing **neglecting** (decreased internal cohesion)
- If  $\Delta\ell_k = 0$ , cohesion is **stable**

## Excite/Dismay

**Excite:** Discussion or actions related to the topic that bring joy, happiness, cheer, enthusiasm in the targeted community or actor [13]

**Dismay:** Discussion or actions related to the topic that create worry, sadness, anger, or fear in that targeted community or actor [13]

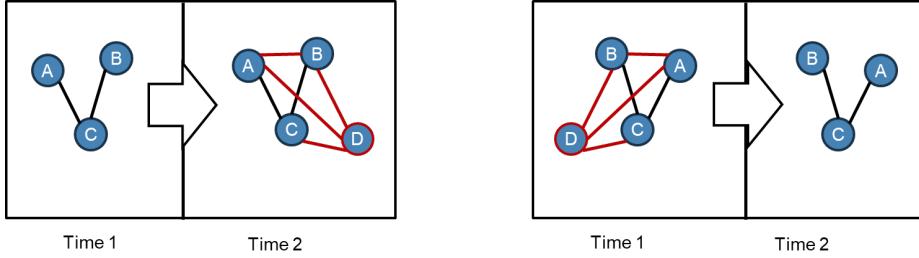


Figure 7.8: Visualization of boost/neglect.

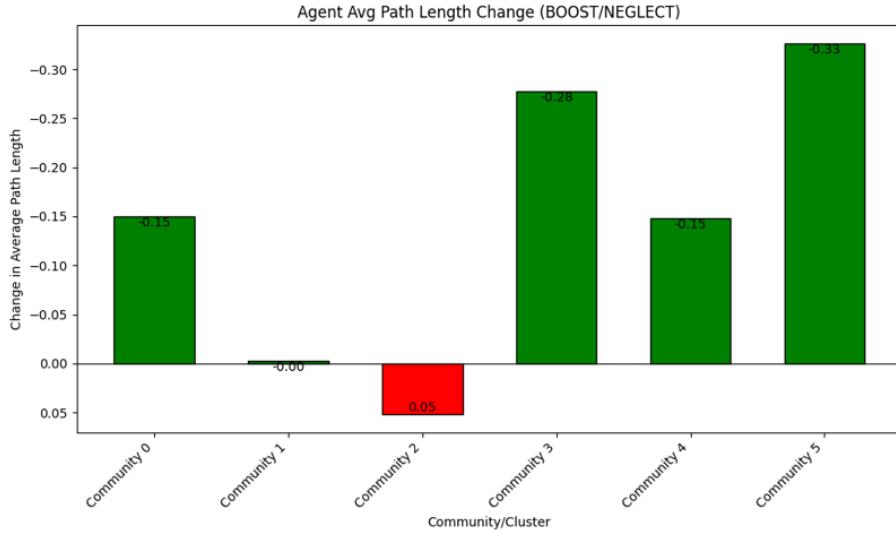


Figure 7.9: Actual results of boost/neglect.

Excite maneuvers include discussion or actions related to a community or topic that cause the reader to experience a positive emotion such as joy, happiness, liking, or excitement. For excite maneuvers, I will look for target output message emotional valence to be higher in happiness over time.

Dismay maneuvers involve discussion or actions related to a community or topic that cause the reader to experience a negative emotion such as worry, sadness, disliking, anger, despair, or fear. As the inverse of excite, I will look for target message emotional valence increasing in anger, sadness, fear over time.

To calculate this, I let  $C_k$  be the tweets in a topic-oriented community, and let  $t_1$  and  $t_2$  denote two time periods.

Define:

- $P_k^{(t)}$ : total number of **positive cues** in cluster  $C_k$  at time  $t$
- $N_k^{(t)}$ : total number of **negative cues** in cluster  $C_k$  at time  $t$
- $T_k^{(t)} = P_k^{(t)} + N_k^{(t)}$ : total sentiment cues in  $C_k$  at time  $t$
- $R_k^{+(t)} = \frac{P_k^{(t)}}{T_k^{(t)}}$ : positive cue ratio

- $R_k^{-(t)} = \frac{N_k^{(t)}}{T_k^{(t)}}$ : negative cue ratio

Then compute the change in positive and negative sentiment ratios:

$$\Delta R_k^+ = R_k^{+(t_2)} - R_k^{+(t_1)}$$

$$\Delta R_k^- = R_k^{-(t_2)} - R_k^{-(t_1)}$$

Interpretation:

- If  $\Delta R_k^+ > 0$ , the cluster is experiencing **Excite**
- If  $\Delta R_k^- > 0$ , the cluster is experiencing **Dismay**
- If both deltas are close to zero, the emotional state is **stable**

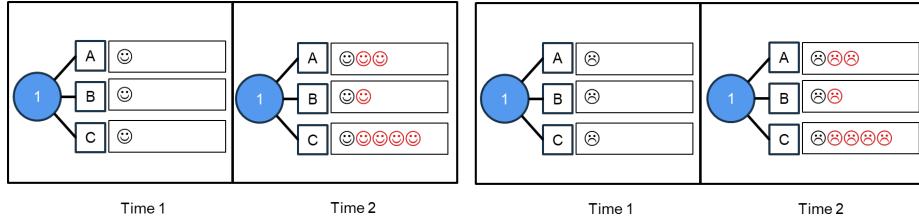


Figure 7.10: Visualization of excite/dismay.

## Explain/Distort

**Explain:** Discussion or actions that provide details on, or elaborate on, a topic to the target community or actor [13]

**Distort:** Discussion or actions that provide unsupportive material that slant the topic for the targeted community or actor [13]

Explain maneuvers will exhibit discussion or actions that clarify a topic to the targeted community or actor often by providing details on, or elaborations on, the topic. For effects-based detection, explain and distort are similar to boost and neglect in the agent space. Explains will decrease the average path length of the concept  $\times$  concept network of cluster - the concepts are more cohesive. A distort will increase the average path length of the concept  $\times$  concept network - the concepts are less cohesive.

So to calculate explain and distort, I let  $C_k$  be a community of concept nodes identified in both time periods  $t_1$  and  $t_2$ , and let  $G_k^{(t)}$  be the subgraph induced by  $C_k$  at time  $t$ , using only internal links (i.e., the concept  $\times$  concept network restricted to nodes in  $C_k$ ).

Define:

- $\ell_k^{(t)}$ : the average shortest path length in  $G_k^{(t)}$

Then compute:

$$\Delta \ell_k = \ell_k^{(t_2)} - \ell_k^{(t_1)}$$

Interpretation:

- If  $\Delta\ell_k < 0$ , then group  $C_k$  is experiencing **explaining** (increased internal cohesion)
- If  $\Delta\ell_k > 0$ , then group  $C_k$  is experiencing **distorting** (decreased internal cohesion)
- If  $\Delta\ell_k = 0$ , cohesion is **stable**

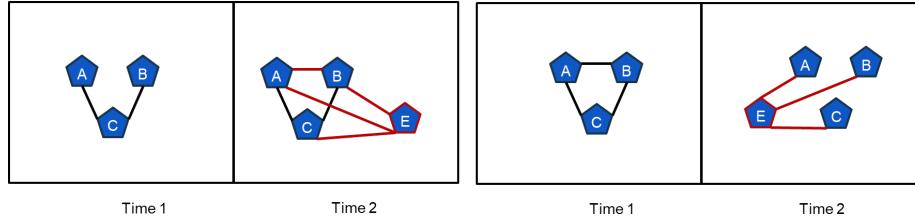


Figure 7.11: Visualization of explain/distort.

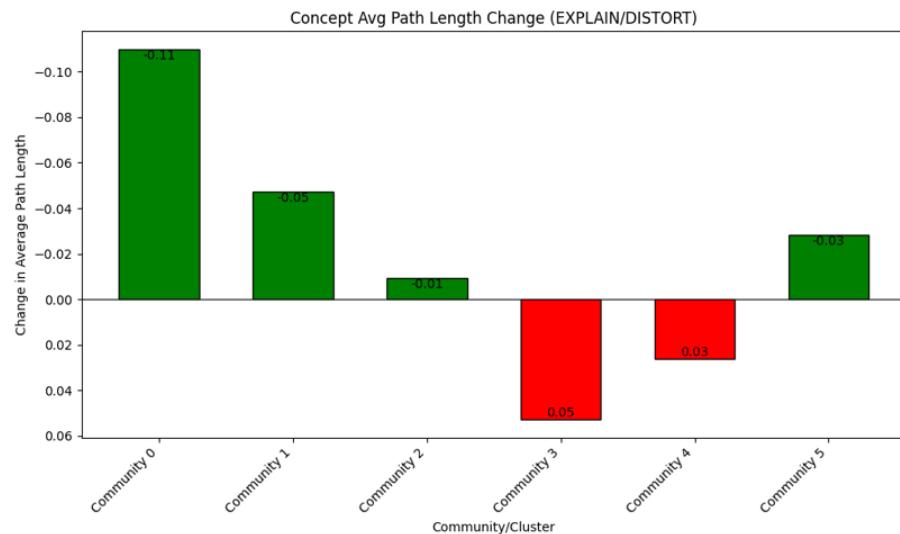


Figure 7.12: Actual results of explain/distort.

## Engage/Dismiss

**Engage:** Discussion or actions that create a personal affinity between the target community or actor and the topic [13]

**Dismiss:** Discussion or actions that suggest that the topic is not important to the targeted community or actor [13]

Engage maneuvers involve discussion or actions that increase the relevance of the topic to the reader often by providing anecdotes or enabling direct participation and so suggesting that the reader can impact the topic or will be impacted by it. In order to detect the effects of engage maneuvers, I will look for a positive change over time in the proportional representation of defining topics within the group and for increased usage of 1st person pronouns. If there is

a decrease in defining topic usage and a decrease in 1st person pronounces, then the group is experiencing the effects of a dismiss.

To calculate engage and dismiss, I let  $C_k$  be a set of messages associated with a group through time, and let  $t_1$  and  $t_2$  denote two time periods.

Let:

- $\mathcal{M}_k^{(t)}$ : the set of messages associated with group  $C_k$  at time  $t$
- $\mathcal{V}$ : the set of all concepts (e.g., hashtags, URLs)
- $f_v^{(t)}$ : the total frequency of concept  $v \in \mathcal{V}$  in  $\mathcal{M}_k^{(t)}$
- $F^{(t)} = \sum_{v \in \mathcal{V}} f_v^{(t)}$ : total concept occurrences in  $\mathcal{M}_k^{(t)}$
- $p_v^{(t)} = \frac{f_v^{(t)}}{F^{(t)}}$ : proportional usage of concept  $v$  in time  $t$
- $\mathcal{V}_{\cap} = \mathcal{V}^{(t_1)} \cap \mathcal{V}^{(t_2)}$ : concepts used in both  $t_1$  and  $t_2$

Also, let:

- $u_1^{(t)}$ : the number of 1st-person pronoun cues in  $\mathcal{M}_k^{(t)}$
- $n^{(t)}$ : the number of messages in  $\mathcal{M}_k^{(t)}$
- $r_1^{(t)} = \frac{u_1^{(t)}}{n^{(t)}}$ : average 1st-person usage per message

Now define:

$$\Delta p = \frac{1}{|\mathcal{V}_{\cap}|} \sum_{v \in \mathcal{V}_{\cap}} (p_v^{(t_2)} - p_v^{(t_1)}) , \quad \Delta r_1 = r_1^{(t_2)} - r_1^{(t_1)}$$

### Interpretation:

$$\begin{cases} \text{Engage} & \text{if } \Delta p > 0 \text{ and } \Delta r_1 > 0 \\ \text{Dismiss} & \text{if } \Delta p < 0 \text{ and } \Delta r_1 < 0 \\ \text{None} & \text{otherwise} \end{cases}$$

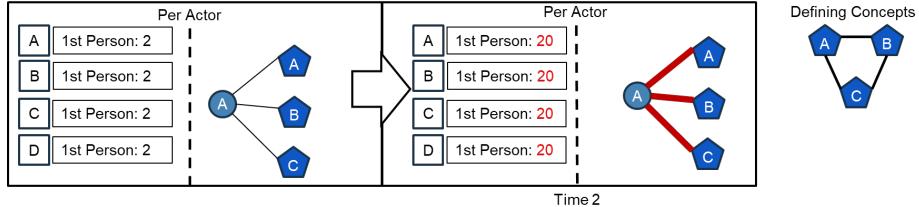


Figure 7.13: Visualization of engage/dismiss.

### Enhance/Distract

**Enhance:** Discussion or actions that provide supportive material that expands the topic for the targeted community or actor [13]

**Distract:** Discussion or actions that redirect the targeted community or actor to a different topic [13]

Enhance maneuvers show discussion or actions that provide material that expands the scope of the topic for the targeted community or actor often by making the topic the master topic to which other topics are linked. To detect the effects of enhance maneuvers I will look for changes in the proportional usage of defining concepts - those concepts used in both time 1 and time 2 - and for distract I will look for a decrease in proportional usage.

To calculate this, I let  $C_k$  be a community tracked across two time periods  $t_1$  and  $t_2$ . Let:

- $\mathcal{M}_k^{(t)}$ : set of messages in cluster  $C_k$  at time  $t$
- $\mathcal{V}_k^{(t)}$ : set of concepts used in  $\mathcal{M}_k^{(t)}$
- $f_v^{(t)}$ : frequency of concept  $v \in \mathcal{V}_k^{(t)}$
- $m_k^{(t)} = |\mathcal{M}_k^{(t)}|$ : number of messages in cluster  $C_k$  at time  $t$

We define the set of concepts used in both time periods as:

$$\mathcal{V}_{\cap} = \mathcal{V}_k^{(t_1)} \cap \mathcal{V}_k^{(t_2)}$$

Let the normalized usage of concept  $v \in \mathcal{V}_{\cap}$  be:

$$p_v^{(t)} = \frac{f_v^{(t)}}{m_k^{(t)}}$$

Then define the \*\*aggregate proportional change\*\* across shared concepts:

$$\Delta p_k = \frac{1}{|\mathcal{V}_{\cap}|} \sum_{v \in \mathcal{V}_{\cap}} (p_v^{(t_2)} - p_v^{(t_1)})$$

### Interpretation:

$$\begin{cases} \text{Enhance} & \text{if } \Delta p_k > 0 \\ \text{Distract} & \text{if } \Delta p_k < 0 \\ \text{Stable} & \text{if } \Delta p_k = 0 \end{cases}$$

The top 10 concepts by magnitude of change (either positive or negative) from each time period are presented for context but are not used in the classification.

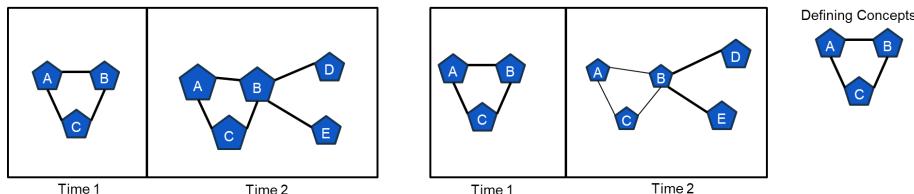


Figure 7.14: Visualization of enhance/distract.

A summary of the maneuvers - their definitions and effects-based metrics can be found in Fig. 7.16 and Fig. 7.17.

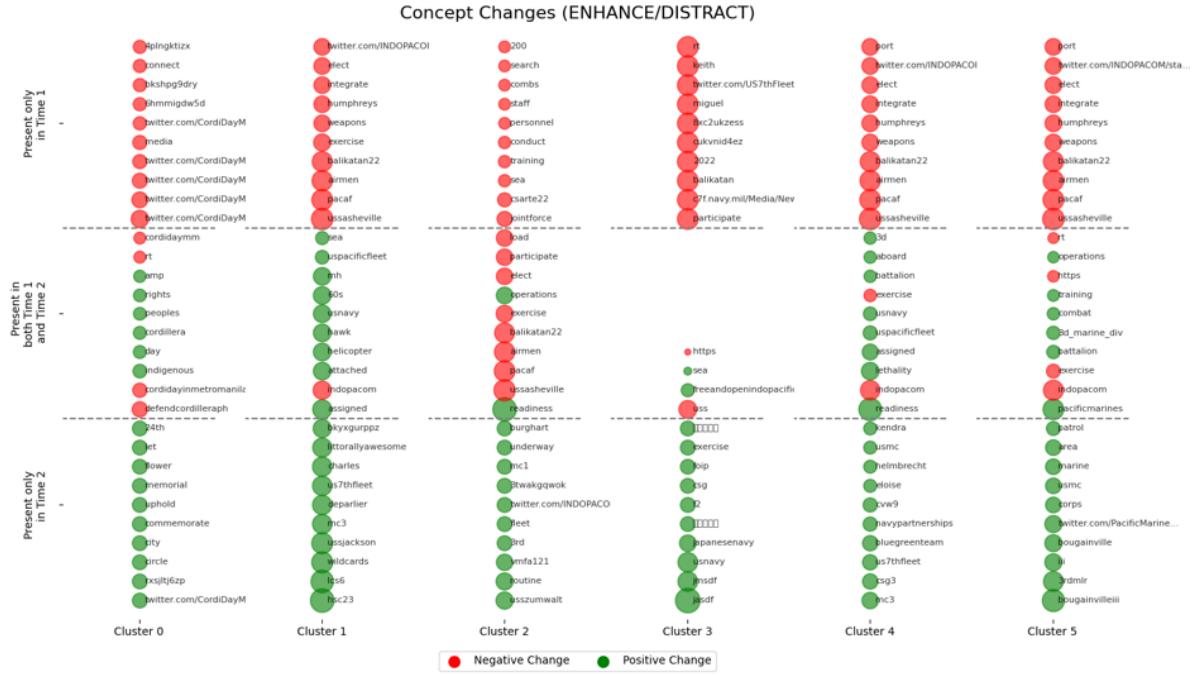


Figure 7.15: Actual results of enhance/distract. Cluster 1 is experiencing an enhance - the time 1+2 concepts are increasing (green) in proportional usage. Cluster 2 is experiencing a distract - proportional usage is decreasing (red)

### 7.3 Temporal Community Detection via Network Intersection and Clustering

In order to accomplish BEND-effects analysis, I have to first establish topic-oriented clusters and topic-oriented communities. As BEND looks at changes in both community and content, these groups must be cohesive from both a network point of view and from a narrative point of view. Therefore, I am constructing groups from agents that have high structurally modularity with surrounding agents both in the agent-interaction space and in the narrative space. To do this, I use the following steps:

Let  $D = \{d_1, d_2, \dots, d_T\}$  be the set of days in the corpus, where  $T$  is the total number of days.

#### 1. Daily Network Construction:

For each day  $d_t \in D$ :

- Construct the agent-agent message network:

$$G_t^{\text{agent}} = (V_t^{\text{agent}}, E_t^{\text{agent}})$$

where:

Name	Definition	Effects-Based Detection
<b>Back</b>	Discussion or actions that increase the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic	Measures: ORA-Pro Super Spreader, Super Friend, General Influencer and Eigenvector Centrality Network(s): Agent x Agent (Strong Ties)
<b>Negate</b>	Discussion or actions that decrease the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic	
<b>Build</b>	Discussion or actions that create a group, or the appearance of a group, where there was none before	Measures: TOG Cluster agent composition
<b>Narrow</b>	Discussion or actions that lead a group to be, or appear to be, more specialized, and possibly to fission, or appear to fission, into two or more distinct groups	
<b>Bridge</b>	Discussion or actions that build a connection between two or more groups or create the appearance of such a connection	Measures: Meta-node (TOG Community) ties Network(s): Agent x Agent (Strong Ties union Concepts)
<b>Neutralize</b>	Discussion or actions that cause a group to be, or appear to be, no longer of relevance, e.g., because it was dismantled	
<b>Boost</b>	Discussion or actions that increase the size of a group and/or the connections among group members, or the appearance of such	Measures: Average Path Length Network(s): Agent x Agent (Strong Ties)
<b>Neglect</b>	Discussion or actions that decrease the size of a group and/or the connections among group members, or the appearance of such	

Figure 7.16: Narrative definitions and metrics

Name	Definition	Effects-Based Detection
<b>Excite</b>	Discussion or actions related to a community or topic that cause the reader to experience a positive emotion such as joy, happiness, liking, or excitement	Measures: Messages of TOG Cluster agents in t2 and in t1 – evaluate for 4 positive/negative CUEs through NetMapper
<b>Dismay</b>	Discussion or actions related to a community or topic that cause the reader to experience a negative emotion such as worry, sadness, disliking, anger, despair, or fear	
<b>Explain</b>	Discussion or actions that clarify a topic to the targeted community or actor often by providing details on, or elaborations on, the topic	Measures: Average Path Length Network(s): Concept x Concept (Agents in TOG)
<b>Distort</b>	Discussion or actions that obscure a topic to the targeted community or actor often by supporting a particular point of view or calling details into question	
<b>Engage</b>	Discussion or actions that increase the relevance of the topic to the reader often by providing anecdotes or enabling direct participation and so suggesting that the reader can impact the topic or will be impacted by it	Measures: A) 1 <sup>st</sup> Person pronoun usage, 2) Defining Concept usage (Defining Concepts are those concepts found in both time 1 and time 2) Network(s): Messages (NetMapper CUEs), Agent x Concept
<b>Dismiss</b>	Discussion or actions that decrease the relevance of the topic to the reader often by providing stories or information that suggest that the reader cannot impact a topic or be impacted by it	
<b>Enhance</b>	Discussion or actions that provide material that expands the scope of the topic for the targeted community or actor often by making the topic the master topic to which other topics are linked	Measures: Proportional change of concept usage of Defining Concepts (concepts found in both time 1 and time 2) Network(s): Concept x Concept (Agents in TOG)
<b>Distract</b>	Discussion or actions that redirect the targeted community or actor to a different topic often by bring up unrelated topics, and making the original topic just one of many	

Figure 7.17: Network definitions and metrics

- $G_t^{\text{agent}}$  is a graph representing strong relation between agents based on retweets, replies, or quotes on day  $d_t$ .
- $V_t^{\text{agent}}$  is the set of agent nodes on day  $t$ .
- $E_t^{\text{agent}}$  is the set of edges representing agent-to-agent connections based on message interactions.
- Construct the concept-concept message network:

$$G_t^{\text{concept}} = (V_t^{\text{concept}}, E_t^{\text{concept}})$$

where:

- $G_t^{\text{concept}}$  is a graph showing co-occurrence of concepts in the same messages on day  $d_t$ .
- $V_t^{\text{concept}}$  is the set of concept nodes (e.g., hashtags, URLs).
- $E_t^{\text{concept}}$  is the set of edges showing concept-concept relationships through shared message context.

- Compute the intersection (logical AND) of the two networks:

$$G_t^{\text{intersect}} = G_t^{\text{agent}} \cap G_t^{\text{concept}}$$

where:

- $G_t^{\text{intersect}}$  retains only those edges that exist in both networks.
- The edge set of the intersection is:

$$E_t^{\text{intersect}} = E_t^{\text{agent}} \cap E_t^{\text{concept}}$$

meaning that an edge  $(u, v)$  is included in  $G_t^{\text{intersect}}$  if and only if  $(u, v)$  is in both  $E_t^{\text{agent}}$  and  $E_t^{\text{concept}}$ .

## 2. Community Detection via Leiden Algorithm:

Apply the Leiden algorithm to each intersected graph:

$$C_t = \text{Leiden}(G_t^{\text{intersect}})$$

where:

- $C_t = \{c_1^t, c_2^t, \dots, c_{k_t}^t\}$  is the set of clusters (communities) detected in  $G_t^{\text{intersect}}$ .
- $k_t$  is the number of clusters at time  $t$ .
- Each  $c_i^t$  is a set of nodes (agents or concepts) grouped together by structural similarity.

## 3. Temporal Cluster Matching:

For each pair of consecutive days  $(t, t+1)$ , compare clusters across time using the Jaccard similarity index:

$$\text{Jaccard}(c_i^t, c_j^{t+1}) = \frac{|c_i^t \cap c_j^{t+1}|}{|c_i^t \cup c_j^{t+1}|}$$

where:

- $c_i^t$  is a cluster at time  $t$ .
- $c_j^{t+1}$  is a cluster at time  $t + 1$ .
- The numerator counts the number of shared nodes between the two clusters.
- The denominator counts the total number of unique nodes across both clusters.

#### 4. Match Criterion:

A pair of clusters is considered a match through time if their Jaccard index is at least 0.5:

$$\text{Jaccard}(c_i^t, c_j^{t+1}) \geq 0.5$$

## 7.4 Results

I applied the temporal clustering community detection and the subsequent BEND-effects maneuver detection to three different datasets.

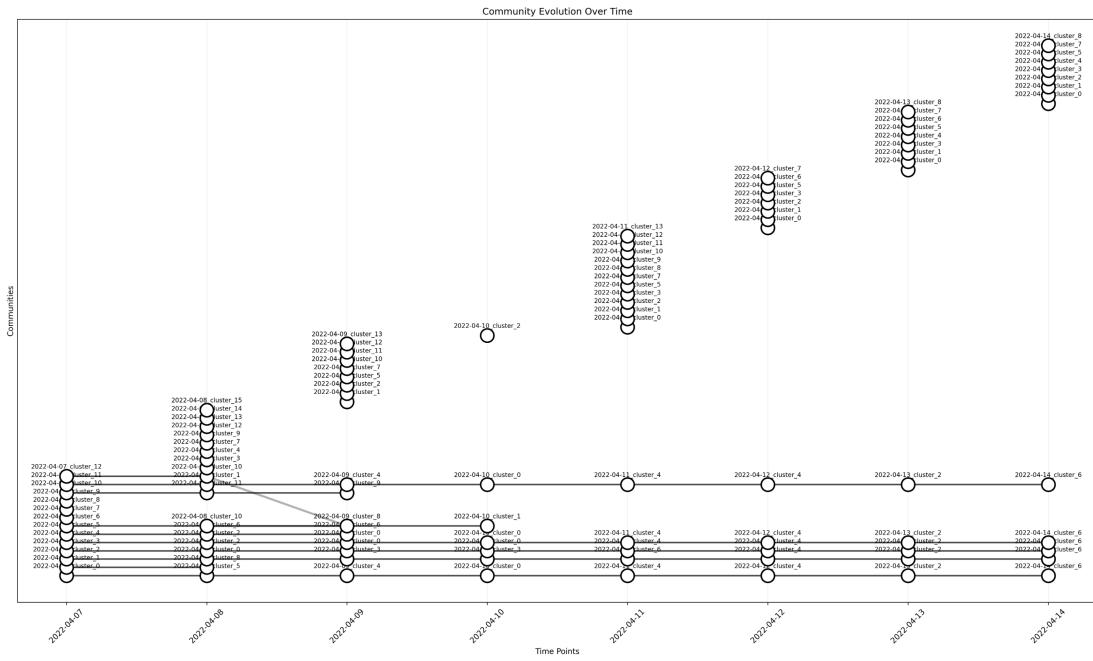


Figure 7.18: Balikatan Groups.

## 7.5 Future Work

BEND-effects is immediately useful for evaluating the effects of social cyber maneuver campaigns within a corpus. However, significant improvements should still be made in the near future.

Each of the maneuver metrics for BEND-effects currently lacks proper thresholding and significance measures. While the existence of maneuver effects is firmly established, their degree

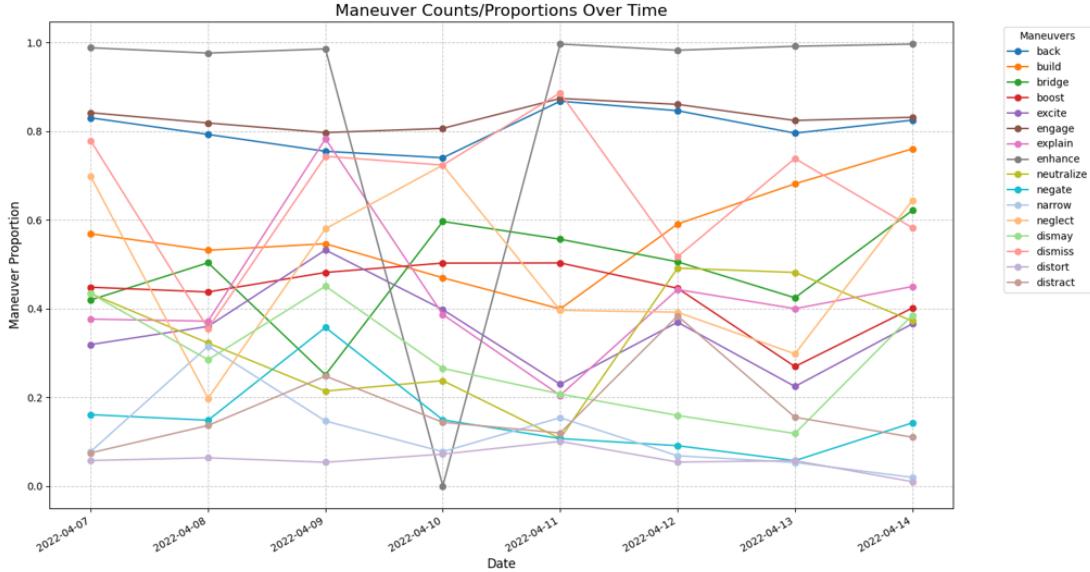


Figure 7.19: Balikatan BEND Counts.

of importance remains left to the interpretation of the analyst. This is the difference between evidence of bridging and evidence of bridging that we care about.

The Jaccard index was used because of performance reasons, but the Adjusted Rand Index (ARI) or Normalized Mutual Information (NMI) methods might produce better results.[41] [70] Jaccard does not account for true negatives - ARI and NMI would - while ARI would help correct for chance and NMI would handle the size changes among clusters better.

The BEND-effects calculations could also be better expressed by first calculating a baseline of the entire corpus at each time period of the given metric and then expressing individual clusters change over time as a residual above and beyond the baseline. This would help ensure the BEND-effect measurement for a cluster is greater than chance and is not a byproduct of a global effect.

## 7.6 Conclusions

BEND-effects is the only automated methodology for evaluating the effectiveness of executed BEND maneuvers and the only comprehensive method for detecting the effects of BEND maneuvers. It relies upon well-established network metrics that align with the contemporary definitions of the BEND maneuvers. Indeed, these network metrics are of such importance that even if BEND is not used as a framework, using the automated detection of the network metrics can serve as a starting point for conducting non-BEND related network and narrative group evaluation.

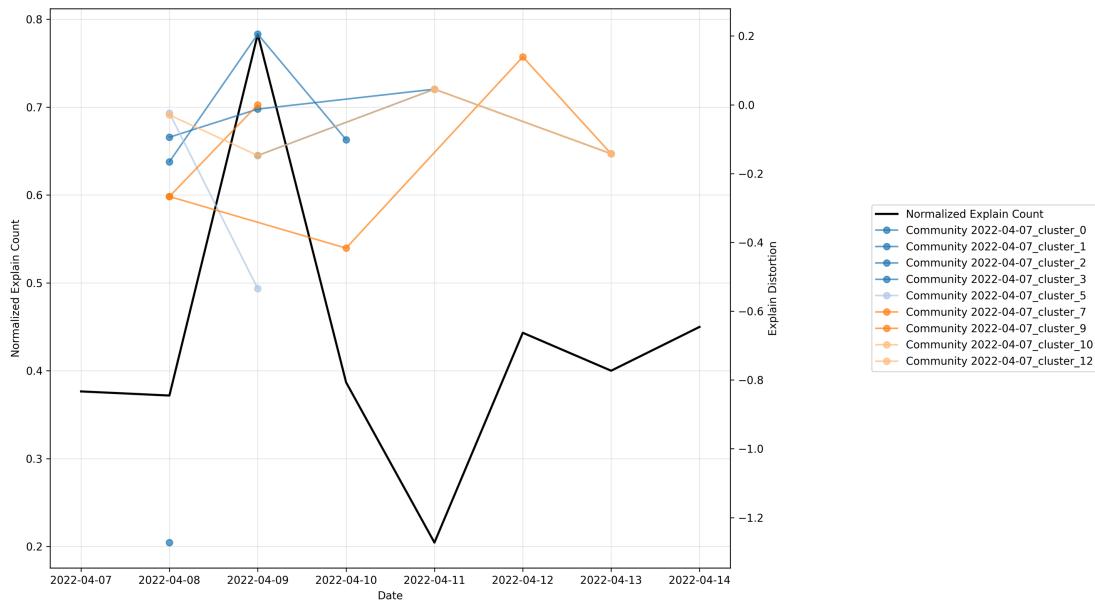


Figure 7.20: Balikatan effect vs maneuver.

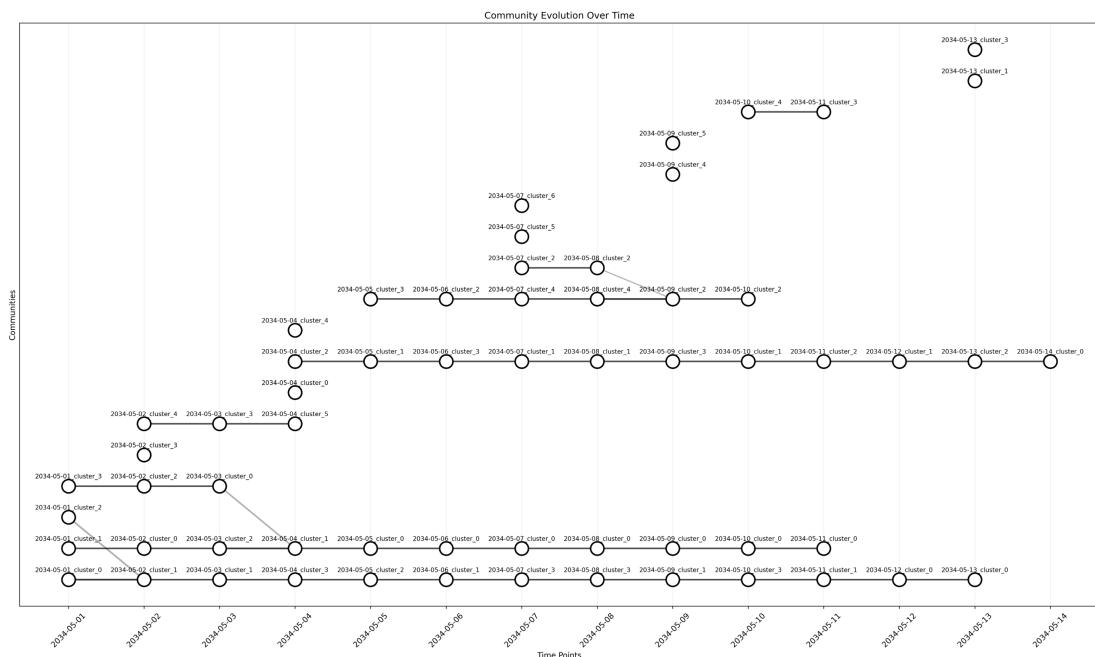


Figure 7.21: ROF Groups.

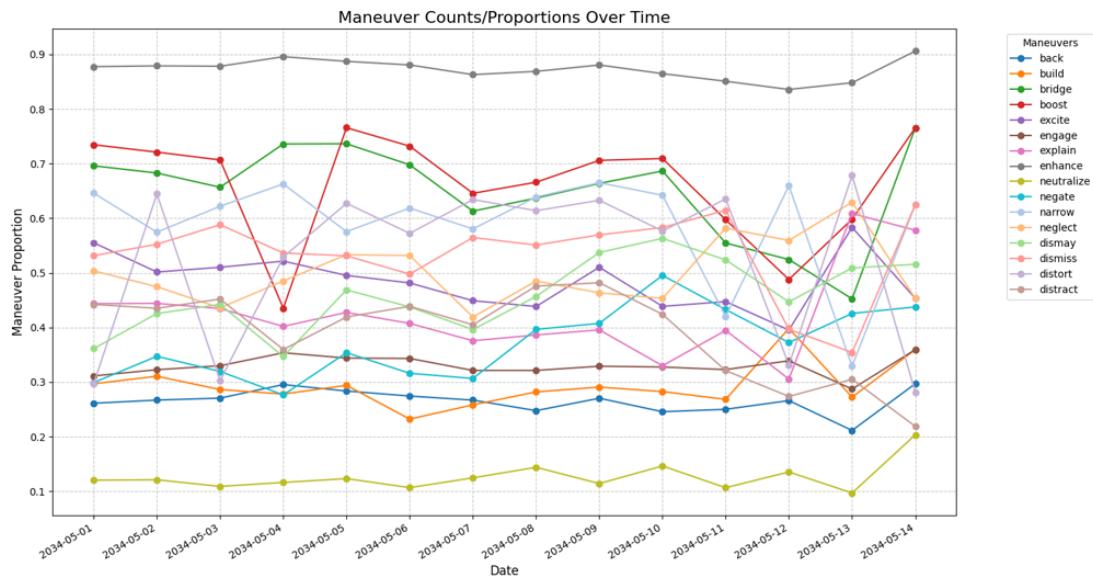


Figure 7.22: ROF BEND Counts.

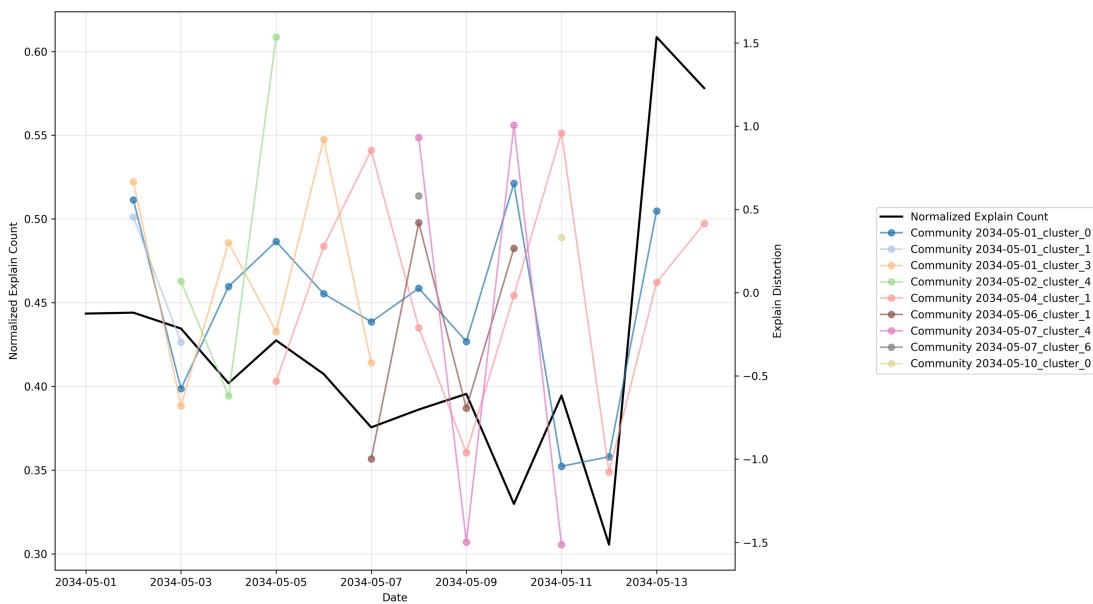


Figure 7.23: ROF effect vs maneuver.

# **Chapter 8**

## **Conclusion**

This dissertation presents a set of novel theoretical and methodological contributions that advance the field of social cybersecurity, particularly in the detection and simulation of influence maneuvers on social media. By integrating military doctrine with computational frameworks and enabling AI-driven scenario generation, this work offers foundational elements for both academic inquiry and applied defense training.

### **8.1 Theoretical Contributions**

A core theoretical contribution of this thesis is the reconceptualization of BEND maneuver detection. While BEND maneuvers have always been effects-based — defined by their impact rather than the intent of their executors — prior detection efforts relied heavily on inferring intent through cues, language, and network signals. This thesis challenges that approach and introduces a framework for effects-based detection of BEND maneuvers. By focusing on observable outcomes rather than inferred motivations, this shift aligns BEND detection with the empirical rigor of other academic and intelligence assessments.

A second theoretical contribution is the refinement of military doctrinal tools to better accommodate social media analysis. Current U.S. military Information Operations doctrine does not apply the same level of analytical precision to social media as it does to other operational environments. This thesis introduces a Social Media MCOO/CSO (Cyber-Social Overlay), providing a structured framework to assess social-cyber terrain and integrate it more effectively into operational planning. This refinement helps bridge the gap between doctrine and the realities of modern information warfare.

### **8.2 Methodological Contributions**

This dissertation also presents several original methodological contributions, advancing both detection capabilities and synthetic scenario generation tools for influence operations.

### **8.2.1 Effects-Based Detection of BEND Maneuvers**

A major methodological advancement of this thesis is the development of an effects-based approach to detecting BEND maneuvers. Unlike previous detection methods that sought to infer intent from content and contextual signals (e.g., CUE+), this approach directly measures the observable effects of a maneuver within a social network or narrative ecosystem. This methodological shift enables:

- More objective and replicable assessments of maneuver effectiveness
- Better integration with automated systems, reducing reliance on subjective human judgment
- Improved post hoc analysis, allowing planners to evaluate whether an observed maneuver actually achieved its intended influence

This framework, in conjunction with CUE+ methods, enables comprehensive detection and empirical assessment of influence maneuvers.

### **8.2.2 SynTel and SynX: Agent-Based Social Media Generators**

This thesis also introduces SynTel and SynX, two agent-based social media generators developed for Telegram and Twitter/X respectively. These tools provide:

- Traditional simulation logic, which models agent behaviors and interactions
- LLM-powered message construction, generating realistic, contextually appropriate content
- Enhanced scenario realism, ensuring that generated training exercises reflect real-world influence dynamics
- Scalable dataset creation, reducing the manual effort required for scenario design

By integrating these capabilities, SynTel and SynX enable the creation of synthetic social media datasets that reflect real-world influence dynamics. These tools provide researchers and practitioners with a controlled yet flexible means of simulating social media influence campaigns in a training or analytical environment.

### **8.2.3 AESOP: AI-Enabled Scenario Orchestration and Planning**

A final methodological innovation is AESOP (AI-Enabled Scenario Orchestration and Planning). AESOP is a planning tool that allows Information Environment planners to:

- Develop social-cyber exercise scenarios from scratch
- Integrate social-cyber vignettes into existing scenarios
- Rapidly generate narrative-driven influence operations training material

## **8.3 Application Contributions**

Beyond its theoretical and methodological impact, this research also contributes practical applications:

- AI-enabled social media scenario development, allowing exercise designers to generate realistic social media narratives from existing datasets or training material
- A draft data standard for social media-based scenario exchange, providing structured interoperability for synthetic data sharing
- Synthetic social media generation for X/Twitter and Telegram, enabling automated creation of training datasets for influence operations exercises

## 8.4 Limitations

**US Military Doctrinal Synthesis** US Information Operations doctrine is evolving and changing rapidly. Many of the referenced Joint and Service Publications are already out of date and the replacement publications are all held at a classified level or have distribution restrictions that prohibit their academic study. Additionally, Information Operations remains a complex issue with authority and titling problems that cannot be resolved in theory and require policy reforms.

**Effects-Based BEND Detection** There is currently no way to directly associate observed BEND effects with any single message BEND maneuver - we are not yet in the precision munitions phase of information environment maneuvers. Also, better methods for measuring BEND maneuvers above baseline are required as residual statistics will be more important than net maneuver counts. Standard ORA-Pro reports do not reflect this need.

**BEND Scenario Development** Without an overarching simulation, training scenario data will be static and unresponsive to training audience feedback. However, AESOP could be used to alter the scenario based upon training audience decisions and new templates could then drive additional synthetic data to get after a highly incremented simulation. Daily static training data is reasonable and appropriate since collection and attribution methods through the social media APIs do not allow for pulling all possible data instantly and continuously.

## Bibliography

- [1] *U. S. Const. amend. IV.* URL <https://www.senate.gov/about/origins-foundations/senate-and-constitution/constitution.htm>. 4.2
- [2] *Privacy Act of 1974, 5 U.S. Code Title 5, section 552a.* URL <https://www.govinfo.gov/content/pkg/USCODE-2018-title5/pdf/USCODE-2018-title5-partI-chap5-subchapII-sec552a.pdf>. 4.2
- [3] *Posse Comitatus Act, U.S. Code Title 18, section 1305.* URL <https://www.law.cornell.edu/uscode/text/18/1385>. 4.2
- [4] *USAF Doctrine Update on Domains and Organizing for Joint Operations.* URL [https://wwwdoctrine.af.mil/Portals/61/documents/doctrine\\_updates/du\\_13\\_09.pdf](https://wwwdoctrine.af.mil/Portals/61/documents/doctrine_updates/du_13_09.pdf). 4.2
- [5] *Summary of the Joint All-Domain Command and Control (JADC2) Strategy.* URL [https://wwwdoctrine.af.mil/Portals/61/documents/doctrine\\_updates/du\\_13\\_09.pdf](https://wwwdoctrine.af.mil/Portals/61/documents/doctrine_updates/du_13_09.pdf). 4.2
- [6] mistralai/mixtral-8x7b-v0.1 · hugging face. URL <https://huggingface.co/mistralai/Mixtral-8x7B-v0.1>. 2.2.1
- [7] Alexander E. Aguilastratt and Matthew S. Updike. The information domain and social media. *Infantry*, 111(1):31–34, 2022. URL [https://www.moore.army.mil/infantry/magazine/issues/2022/Spring/PDF/6\\_ProfessionalForum.pdf](https://www.moore.army.mil/infantry/magazine/issues/2022/Spring/PDF/6_ProfessionalForum.pdf). 4.2
- [8] Alexandre Alaphilippe. Adding a d to the ABC disinformation framework. URL <https://policycommons.net/artifacts/4139892/adding-a-d-to-the-abc-disinformation-framework/4948112/>. Publisher: Brookings Institution. 3.1
- [9] Iuliia Alieva, J. D. Moffitt, and Kathleen M. Carley. How disinformation operations against russian opposition leader alexei navalny influence the international audience on twitter. 12 (1):80. ISSN 1869-5450, 1869-5469. doi: 10.1007/s13278-022-00908-6. URL <https://link.springer.com/10.1007/s13278-022-00908-6>. 3.1
- [10] Adam Badawy, Aseel Addawood, Kristina Lerman, and Emilio Ferrara. Characterizing the 2016 russian IRA influence campaign. 9(1):31. ISSN 1869-5450, 1869-5469. doi: 10.1007/s13278-019-0578-6. URL <http://link.springer.com/10.1007/s13278-019-0578-6>. 1.1
- [11] David M. Beskow and Kathleen M. Carley. Agent based simulation of bot disinformation maneuvers in twitter. In *2019 Winter Simulation Conference (WSC)*, pages 750–761. IEEE, . ISBN 978-1-72813-283-9. doi: 10.1109/WSC40007.2019.9004942. URL <https://ieeexplore.ieee.org/document/9004942/>. 6.3
- [12] David M. Beskow and Kathleen M. Carley. Social cybersecurity: an emerging national security requirement, . URL <https://apps.dtic.mil/sti/citations/AD1108494>. 3.1

- [13] Janice Blane. Social-cyber maneuvers for analyzing online influence operations. page 13152760 Bytes. doi: 10.1184/R1/22825112.V1. URL [https://kilthub.cmu.edu/articles/thesis/Social-Cyber\\_Maneuvers\\_for\\_Analyzing\\_Online\\_Influence\\_Operations/22825112/1](https://kilthub.cmu.edu/articles/thesis/Social-Cyber_Maneuvers_for_Analyzing_Online_Influence_Operations/22825112/1). 2.1, 3.1, 3.1, 7.1, 7.2, 7.2, 7.2, 7.2, 7.2, 7.2, 7.2, 7.2, 7.2, 7.2, 7.2
- [14] Janice T Blane, Daniele Bellutta, and Kathleen M Carley. Social-cyber maneuvers during the COVID-19 vaccine initial rollout: Content analysis of tweets. 24(3):e34040, . ISSN 1438-8871. doi: 10.2196/34040. URL <https://www.jmir.org/2022/3/e34040>. 2.1, 3.1
- [15] Janice T. Blane, J. D. Moffitt, and Kathleen M. Carley. Simulating social-cyber maneuvers to deter disinformation campaigns. In Robert Thomson, Muhammad Nihal Husain, Christopher Dancy, and Aryn Pyke, editors, *Social, Cultural, and Behavioral Modeling*, volume 12720, pages 153–163. Springer International Publishing, . ISBN 978-3-030-80386-5 978-3-030-80387-2. doi: 10.1007/978-3-030-80387-2\_15. URL [https://link.springer.com/10.1007/978-3-030-80387-2\\_15](https://link.springer.com/10.1007/978-3-030-80387-2_15). 6.2
- [16] Sam Blazek. SCOTCH: A framework for rapidly assessing influence operations. URL <https://www.atlanticcouncil.org/blogs/geotech-cues/scotch-a-framework-for-rapidly-assessing-influence-operations/>. 3.1
- [17] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. 3:993–1022. 5.3
- [18] John G. Breslin, Uldis Bojārs, and Stefan Decker. Sioc - semantically-interlinked online communities. <https://web.archive.org/web/20220331224416/http://sioc-project.org/>. Accessed: 2025-03-13. 5.4
- [19] Dan Brickley and Libby Miller. Foaf vocabulary specification 0.99. <http://xmlns.com/foaf/spec/>, 2014. Accessed: 2025-03-13. 5.4
- [20] Kathleen M. Carley. Social cybersecurity: an emerging science. 26(4):365–381. ISSN 1572-9346. doi: 10.1007/s10588-020-09322-9. URL <https://doi.org/10.1007/s10588-020-09322-9>. 3.1
- [21] Kathleen M. Carley, Guido Cervone, Nitin Agarwal, and Huan Liu. Social cybersecurity. In Robert Thomson, Christopher Dancy, Ayaz Hyder, and Halil Bisgin, editors, *Social, Cultural, and Behavioral Modeling*, volume 10899, pages 389–394. Springer International Publishing, . ISBN 978-3-319-93371-9 978-3-319-93372-6. doi: 10.1007/978-3-319-93372-6\_42. URL [https://link.springer.com/10.1007/978-3-319-93372-6\\_42](https://link.springer.com/10.1007/978-3-319-93372-6_42). 3.1
- [22] L Richard Carley, Jeff Reminga, and Kathleen M Carley. Ora & netmapper. In *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation*. Springer, volume 3, page 7, . Issue: 3.3. 2.2.1, 2.2.1, 3.1
- [23] *Joint Information Operations Proponent*. Chairman of the Joint Chiefs of Staff, Washington, DC, USA, 2014. URL <https://www.jcs.mil/Portals/36/Documents/>

- [24] Serina Chang, Alicja Chaszczewicz, Emma Wang, Maya Josifovska, Emma Pierson, and Jure Leskovec. Llms generate structurally realistic social networks but overestimate political homophily, 2024. URL <https://arxiv.org/abs/2408.16629>. These authors contributed equally: Serina Chang, Alicja Chaszczewicz. 6.2, 6.3, 6.5
- [25] Bastien Chopard and Michel Droz. Cellular automata: Modelling of physical systems. In *Cellular Automata Modeling of Physical Systems*, pages 6–13. Cambridge University Press, Cambridge, MA, 1998. URL <https://doi.org/10.1017/CBO9780511549755.6.2>
- [26] Fred Cohen. Simulating cyber attacks, defences, and consequences. 18(6):479–518. ISSN 01674048. doi: 10.1016/S0167-4048(99)80115-1. URL <https://linkinghub.elsevier.com/retrieve/pii/S0167404899801151>. 3.1
- [27] Robert Cordray III and Marc J. Romanych. Mapping the information environment. pages 7–10. URL <https://www.quantico.marines.mil/Portals/147/Docs/MCIOC/IORecruiting/MappingtheInformationEnvironmentIOSphereSummer2005.pdf>. 4.2
- [28] *Information Operations (IO)*. Department of Defense, Washington, DC, USA, 2017. URL <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/360001p.pdf>. 4.2
- [29] *Information in Air Force Operations*. Department of the Air Force, Washington, DC, USA, 2018. URL [https://wwwdoctrine.af.mil/Portals/61/documents/AFDP\\_3-13/3-13-AFDP-INFO-OPS.pdf](https://wwwdoctrine.af.mil/Portals/61/documents/AFDP_3-13/3-13-AFDP-INFO-OPS.pdf). 3.1, 4.2
- [30] *The Operations Process*. Department of the Army, Washington, DC, USA, 2019. URL [https://armypubs.army.mil/epubs/DR\\_pubs/DR\\_a/ARN18126-ADP\\_5-0-000-WEB-3.pdf](https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN18126-ADP_5-0-000-WEB-3.pdf). 4.2
- [31] *Intelligence Preparation of the Battlefield*. Department of the Army, Washington, DC, USA, 2019. URL <https://irp.fas.org/doddir/army/atp2-01-3.pdf>. 3.1, 4.2
- [32] *Army Information*. Department of the Army, Washington, DC, USA, 2023. URL <https://irp.fas.org/doddir/army/adp3-13.pdf>. 4.2
- [33] *Navy Information Operations*. Department of the Navy, Washington, DC, USA, 2014. URL [https://www.usna.edu/Training/\\_files/documents/References/3C%20MQS%20References/2015-2016%203C%20MQS%20References/NWP%203-13\\_Information%20Operations\\_FEB2014.pdf](https://www.usna.edu/Training/_files/documents/References/3C%20MQS%20References/2015-2016%203C%20MQS%20References/NWP%203-13_Information%20Operations_FEB2014.pdf). 4.2
- [34] Stephen Dipple and Kathleen M. Carley. *Construct User Guide*. URL <https://www.cmu.edu/casos-center/publications/cmu-s3d-23-104.pdf>. 6.2
- [35] Geoffrey B. Dobson and Kathleen M. Carley. Cyber-FIT: An agent-based modelling approach to simulating cyber warfare. In Dongwon Lee, Yu-Ru Lin, Nathaniel Osgood, and Robert Thomson, editors, *Social, Cultural, and Behavioral Modeling*, volume 10354, pages 139–148. Springer International Publishing. ISBN 978-3-319-60239-4 978-3-319-60240-

0. doi: 10.1007/978-3-319-60240-0\_18. URL [https://link.springer.com/10.1007/978-3-319-60240-0\\_18](https://link.springer.com/10.1007/978-3-319-60240-0_18). 3.1
- [36] Netherlands General Intelligence Federal Bureau of Investigation (FBI), Cyber National Mission Force (CMNF), Netherlands Military Intelligence Security Service (AIVD), the Netherlands Police (DNP) Security Service (MIVD), and the Canadian Center for Cyber Security (CCCS). State-sponsored russian media leverages meliorator software for foreign malign influence activity. Technical Report CSA-2024-0709, Federal Bureau of Investigation (FBI), Cyber National Mission Force (CMNF), Netherlands General Intelligence and Security Service (AIVD), Netherlands Military Intelligence and Security Service (MIVD), the Netherlands Police (DNP), and the Canadian Center for Cyber Security (CCCS), July 2024. URL <https://www.ic3.gov/CSA/2024/240709.pdf>. Joint Cybersecurity Advisory. 5.4
- [37] Camille François. Actors, behaviors, content: A disinformation ABC. URL [http://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/05/ABC\\_Framework\\_TWG\\_Francois\\_Sept\\_2019.pdf](http://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/05/ABC_Framework_TWG_Francois_Sept_2019.pdf). 3.1
- [38] Xu Guo and Yiqiang Chen. Generative ai for synthetic data generation: Methods, challenges and the future. *arXiv preprint arXiv:2403.04190*, 2024. URL <https://arxiv.org/pdf/2403.04190v1.pdf>. 5.1, 6.2
- [39] Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. Evaluating large language models in generating synthetic hci research data: a case study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394215. doi: 10.1145/3544548.3580688. URL <https://doi.org/10.1145/3544548.3580688>. 6.2
- [40] *Marine Air-Ground Task Force Information Operations*. Headquarters United States Marine Corps, Department of the Navy, Washington, DC, USA, 2018. URL <https://www.marines.mil/Portals/1/Publications/MCWP%203-32.pdf>. 4.2
- [41] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985. doi: 10.1007/BF01908075. 7.5
- [42] Charity S. Jacobs, Lynnette Hui Xian Ng, and Kathleen M. Carley. Tracking china's cross-strait bot networks against taiwan. In *Social, Cultural, and Behavioral Modeling (SBP-BRIMS 2023)*, pages 115–125, New York, NY, 2023. Springer. doi: 10.1007/978-3-031-43129-6\_12. URL [https://link.springer.com/chapter/10.1007/978-3-031-43129-6\\_12](https://link.springer.com/chapter/10.1007/978-3-031-43129-6_12). 6.4
- [43] *Joint Targeting*. Joint Chiefs of Staff, Washington, DC, USA, 2013. URL [https://jfsc.ndu.edu/Portals/72/Documents/JC2IOS/Additional\\_Reading/1F4\\_jp3-60.pdf](https://jfsc.ndu.edu/Portals/72/Documents/JC2IOS/Additional_Reading/1F4_jp3-60.pdf). 4.2, 4.2
- [44] *Joint Intelligence Preparation of the Operational Environment*. Joint Chiefs of Staff, Washington, DC, USA, 2014. URL <https://irp.fas.org/doddir/dod/jp2-01-3.pdf>. 4.2, 4.2
- [45] *Information Operations*. Joint Chiefs of Staff, Washington, DC, USA, 2014. URL <https://>

- //irp.fas.org/doddir/dod/jp3\_13.pdf. 3.1, 4.2
- [46] *Public Affairs*. Joint Chiefs of Staff, Washington, DC, USA, 2016. URL [https://irp.fas.org/doddir/dod/jp3\\_61.pdf](https://irp.fas.org/doddir/dod/jp3_61.pdf). 4.2
- [47] *Doctrine for the Armed Forces of the United States*. Joint Chiefs of Staff, Washington, DC, USA, 2017. URL <https://fas.org/irp/doddir/dod/jp1.pdf>. 4.2
- [48] *Joint Operations*. Joint Chiefs of Staff, Washington, DC, USA, 2017. URL [https://irp.fas.org/doddir/dod/jp3\\_0.pdf](https://irp.fas.org/doddir/dod/jp3_0.pdf). 4.2
- [49] Jeon-Hyung Kang and Kristina Lerman. LA-CTR: A limited attention collaborative topic regression for social media. 27(1):1128–1134. ISSN 2374-3468, 2159-5399. doi: 10.1609/aaai.v27i1.8451. URL <https://ojs.aaai.org/index.php/AAAI/article/view/8451>. 6.3
- [50] Catherine King, Christine Sowa Lepird, and Kathleen M. Carley. Project OMEN: Designing a training game to fight misinformation on social media. 3. URL <http://reports-archive.adm.cs.cmu.edu/anon/anon/usr0/ftp/home/ftp/isr2021/CMU-ISR-21-110.pdf>. 4.2, 5.2
- [51] Santosh Kulkarni, Joyanta Banerjee, and Mallik Panchumarthy. Generate synthetic counterparty (cr) risk data with generative ai using amazon bedrock llms and rag, February 2025. URL [https://aws.amazon.com/blogs/machine-learning/generate-synthetic-counterparty-cr-risk-data-with-generative-ai-using-an-amazon-web-services-machine-learning-blog](https://aws.amazon.com/blogs/machine-learning/generate-synthetic-counterparty-cr-risk-data-with-generative-ai-using-an-amazon-web-services-machine-learning-blog/). 6.2
- [52] Christine Sowa Lepird. *Pink Slime: Measuring, Finding, and Countering Online Threats to Local News*. PhD thesis, Carnegie Mellon University, October 2024. URL [https://christine-lepird.github.io/Lepird\\_Thesis\\_Oct2024.pdf](https://christine-lepird.github.io/Lepird_Thesis_Oct2024.pdf). 6.5, 6.3, 6.5.1
- [53] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Namnan Goyal, Heinrich Küttler, Nils Reimers, Mohammad Shoeybi, Mostafa Dehghani, Sébastien Borgeaud, Sean Welleck, and Sebastian Riedel. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 9459–9474, 2020. URL <https://arxiv.org/abs/2005.11401>. 5.5
- [54] Yao Lu, Peng Zhang, Yanan Cao, Yue Hu, and Li Guo. On the frequency distribution of retweets. 31:747–753. ISSN 18770509. doi: 10.1016/j.procs.2014.05.323. URL <https://linkinghub.elsevier.com/retrieve/pii/S1877050914005006>. 6.3
- [55] Dennis L. Meadows, William W. Behrens, Donella H. Meadows, Roger F. Naill, Jørgen Randers, and Erich Zahn. *Dynamics of Growth in a Finite World*. Wright-Allen Press, Cambridge, MA, 1974. URL [https://elmoukrie.com/wp-content/uploads/2020/12/dynamics-of-growth-in-a-finite-world\\_nodrm.pdf](https://elmoukrie.com/wp-content/uploads/2020/12/dynamics-of-growth-in-a-finite-world_nodrm.pdf). 6.2
- [56] Randall Munroe. Standards. <https://xkcd.com/927/>, 2011. Accessed: 2025-03-13. 5.4
- [57] Lynnette Hui Xian Ng and Kathleen M. Carley. Deflating the chinese balloon: types of

- twitter bots in us-china balloon incident. *EPJ Data Science*, 12(1):63, 2023. doi: 10.1140/epjds/s13688-023-00440-3. URL <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-023-00440-3>. 6.4
- [58] Ben Nimmo. Anatomy of an info-war: How russia's propaganda machine works, and how to counter it. 15:1–16. URL <https://www.stopfake.org/en/anatomy-of-an-info-war-how-russia-s-propaganda-machine-works-and-how-to-counter-it>. 3.1
- [59] OASIS Open. Cti documentation. <https://oasis-open.github.io/cti-documentation/>. Accessed: 2025-03-13. 5.4
- [60] Office of the Director of National Intelligence. Information resource metadata. <https://www.dni.gov/index.php/who-we-are/organizations/ic-cio/ic-technical-specifications/information-resource-metadata>. Accessed: 2025-03-13. 5.4
- [61] oobabooga. A gradio web UI for large language models: oobabooga/text-generation-webui. URL <https://github.com/oobabooga/text-generation-webui>. original-date: 2022-12-21T04:17:37Z. 5.3
- [62] *The official Python library for the OpenAI API*. OpeanAI. URL <https://github.com/openai/openai-python>. original-date: 2020-10-25T23:23:54Z. 2.2.1, 5.3
- [63] James Pamment. *The EU's Role in Fighting Disinformation: Crafting A Disinformation Framework*. Carnegie Endowment for International Peace. URL [https://carnegieendowment.org/files/Pamment\\_-\\_Crafting\\_Disinformation\\_1.pdf](https://carnegieendowment.org/files/Pamment_-_Crafting_Disinformation_1.pdf). 3.1
- [64] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior, 2023. URL <https://arxiv.org/pdf/2304.03442.pdf>. 6.2
- [65] Cesar Augusto Rodriguez, Timothy C. Walton, and Chu Hyong. Putting the “FIL” into “DIME”: growing joint understanding of the instruments of power. 97(8). URL <https://apps.dtic.mil/sti/citations/tr/AD1099537>. 4.2
- [66] James Snell, Evan Prodromou, and Sarven Capadisli. Activitystreams 2.0 core vocabulary. <https://www.w3.org/TR/activitystreams-core/>, 2017. W3C Recommendation, Accessed: 2025-03-13. 5.4
- [67] Qt for Python Team. PySide6: Python bindings for the qt cross-platform application and UI framework. URL <https://pyside.org>. 2.2.2, 5.3
- [68] Joshua Uyheng, Thomas Magelinski, Ramon Villa-Cox, Christine Sowa, and Kathleen M. Carley. Interoperable pipelines for social cyber-security: assessing twitter information operations during NATO trident juncture 2018. 26(4):465–483. ISSN 1381-298X, 1572-9346. doi: 10.1007/s10588-019-09298-1. URL <http://link.springer.com/10.1007/s10588-019-09298-1>. 3.1
- [69] Veniamin Veselovsky, Manoel Horta Ribeiro, Akhil Arora, Martin Josifoski, Ashton Anderson, and Robert West. Generating faithful synthetic data with large language models: A

- case study in computational social science. *arXiv preprint arXiv:2305.15041*, 2023. URL <https://arxiv.org/pdf/2305.15041.pdf>. 6.2
- [70] Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. In *Journal of Machine Learning Research*, volume 11, pages 2837–2854, 2010. URL <https://jmlr.csail.mit.edu/papers/volume11/vinh10a/vinh10a.pdf>. 7.5
- [71] Kritin Vongthongsri. Using llms for synthetic data generation: The definitive guide. <https://www.confident-ai.com/blog/the-definitive-guide-to-synthetic-data-generation-using-llms>, 2023. Confident AI Blog, Accessed: 2025-03-14. 6.2
- [72] X. X ids overview, 2023. URL <https://docs.x.com/resources/fundamentals/x-ids>. Describes the structure and generation of Snowflake-based unique identifiers used by X (formerly Twitter). 5.3

# **Appendix A**

## **AESOP Products**

The following are raw outputs from AESOP based on a synthetic scenario.

# Scenario Summary

## SYNTHETIC\_SCENARIO

---

### General Info

**Date Range:** 2034-05-01 to 2034-05-14

**Countries of Interest:**

United States

**Regions of Interest:**

South-eastern Asia

**Other Locations of Interest:**

**Primary Training Audience:**

JTF Information Operations Coordination Cell

**Primary Training Audience's Higher Headquarters' Mission:**

To deter aggression, enhance regional stability, and safeguard US and allied interests within the Pacific theater, the JTF will conduct integrated military operations in response to any emerging crises. The JTF will prioritize defense of key strategic partners, ensure freedom of navigation, and enhance interoperability among US forces and regional allies to preserve security and stability across the Indo-Pacific

**Primary Training Audience's Mission:**

To plan, synchronize, and execute information operations that support Joint Task Force objectives by shaping the information environment, countering adversary disinformation, and reinforcing US and allied narratives. The IOCC will ensure timely dissemination of truthful information to influence key audiences, protect friendly information systems, and degrade adversarial information campaigns across the Pacific theater. The cell will focus on countering disinformation related to ongoing crisis, strengthening US-ally relations during exercise, and mitigating the influence of adversaries in the Pacific.

**Strategic Communication Guidance and Objectives:**

Strategic Communication Guidance

o Reinforce US Commitment to Regional Allies and Partners: Emphasize USINDOPACOM's commitment to defending allies. Highlight the US as a steadfast partner, ready to uphold international law and stability in the region.

o Promote Legitimacy and Transparency: Maintain credibility by providing truthful, transparent, and consistent communication. Ensure that all messaging adheres to international norms and reinforces the legitimate nature of US and allied actions, particularly in responding to humanitarian crises and protecting regional sovereignty.

EXERCISE // UNCLASSIFIED // EXERCISE  
All material contained is confidential and proprietary  
to Netanomics and the CASOS center at CMU.

- o Counter Adversary Disinformation: Actively identify, expose, and counter false narratives from adversaries. Use both defensive (disproving false claims) and offensive (undermining the adversary's credibility) communication strategies.
- o Amplify Humanitarian Assistance and Disaster Relief (HADR) Efforts: Leverage the JTF's response to disasters to highlight US and allied humanitarian efforts, enhancing regional perception of the US as a force for good and stability. Showcase coordination with allies
- o Strengthen Public Confidence in Exercises: Publicize the importance of US-ally cooperation through exercises, emphasizing the role of these exercises in improving readiness, interoperability, and regional defense capabilities.
- o Focus on Shared Values and Collective Security: Highlight shared values such as democracy, sovereignty, and freedom of navigation. Underscore the importance of collective security arrangements in preserving peace and stability in the Indo-Pacific region.
- o Enhance Operational Security (OPSEC): Ensure that messaging balances the need for transparency with the need to protect sensitive operational details. Avoid revealing critical information that could be exploited by adversaries.

#### Strategic Communication Objectives

- o Deter Aggression: Use strategic messaging to signal the US's resolve and capability to defend allies and other regional interests, creating uncertainty in the minds of adversaries about the costs of aggression. The objective is to prevent escalation by demonstrating readiness and resolve.
- o Undermine Adversary Morale and Unity: Undermine adversary military cohesion and legitimacy in joint exercises by exposing potential weaknesses, inconsistencies in messaging, or highlighting tensions within their alliance. Promote doubt in the efficacy of their operations and decision-making processes.
- o Influence Regional and Global Audiences: Shape perceptions among regional allies, neutral states, and global audiences by framing the JTF's actions as defensive and stabilizing. Encourage regional actors to align with US and allied positions rather than adopting a neutral or adversarial stance.
- o Promote US and Allied Strength and Unity: Strengthen the perception of US military strength, technological superiority, and the unity of the JTF and its allies. Convey a sense of overwhelming capability and preparedness to respond to any crisis.
- o Build Trust in US-Led Humanitarian and Stability Operations: Build trust in US-led disaster relief operations by showcasing swift and effective humanitarian assistance.

EXERCISE // UNCLASSIFIED // EXERCISE  
 All material contained is confidential and proprietary  
 to Netanomics and the CASOS center at CMU.

Increase confidence in US leadership and response capabilities during natural disasters and crises.

- o Support Public Opinion and Morale in the US and Allied Nations: Ensure domestic and allied public support for the JTF's operations by emphasizing the necessity of US presence and actions in maintaining regional stability, protecting trade routes, and upholding international law.
- o Expose Adversary Malign Activities: Publicize any malign activities or violations of international law by adversaries, such as posturing in joint exercises or disinformation campaigns, to paint these adversaries as destabilizing forces in the region.

#### Key Messaging Themes

- o Peace and Stability: US and allied actions are aimed at preserving peace and maintaining regional stability.
- o Humanitarian Response: Highlight the US's role as a provider of humanitarian aid, particularly in Japan.
- o International Norms and Rules: Emphasize adherence to international law and the importance of sovereignty and self-determination for regional partners.
- o US Resolve and Strength: Underscore the capability and commitment of US forces to defend allies and protect shared interests.
- o Collective Security: Promote regional cooperation and multilateral partnerships as essential for maintaining stability against shared threats.

EXERCISE // UNCLASSIFIED // EXERCISE  
All material contained is confidential and proprietary  
to Netanomics and the CASOS center at CMU.

# Actor/Organization Example

## United Nations

---



### General Info

Type: Government

Organizational Leader: Anna Virtanen

Main Account Handle: UN

Other Associated Handles:

Handle:

Screen Name:

URL:

#### Description:

The United Nations acts as a global forum for addressing pressing international issues and promoting dialogue and cooperation among nations. It provides a platform for member states to work together on matters such as socioeconomic development, human rights, climate change, peacekeeping, and disarmament. The organization is guided by the principles of sovereign equality, impartiality, and non-interference in the domestic affairs of its member states.

Throughout its history, the UN has played a significant role in fostering international peace and stability. One of its earliest successes came with the establishment of the United Nations Partition Plan for Palestine in 1947, which aimed to address the ongoing conflict in the region. Over the years, the UN has continued to deploy peacekeeping missions to various parts of the world, including in Cyprus, the Democratic Republic of Congo, and South Sudan, among others.

EXERCISE // UNCLASSIFIED // EXERCISE  
All material contained is confidential and proprietary  
to Netanomics and the CASOS center at CMU.

The UN has also been instrumental in promoting human rights and advancing global development. It has been at the forefront of efforts to combat poverty, hunger, and disease, and has spearheaded initiatives such as the Millennium Development Goals and the Sustainable Development Goals. Additionally, the organization has played a pivotal role in addressing global health crises, such as the HIV/AIDS epidemic and the recent COVID-19 pandemic.

The UN is composed of several principal organs, including the General Assembly, the Security Council, the Economic and Social Council, the Secretariat, and the International Court of Justice. These organs work together to carry out the organization's mandate, with member states contributing to decision-making through their participation in the General Assembly and the Security Council.

Overall, the United Nations continues to serve as a vital platform for international cooperation and diplomacy. As the global challenges facing our world become increasingly complex, the UN plays an essential role in fostering dialogue, promoting collaboration, and working towards a more peaceful and equitable world for all.

# Actor/Person Example

## Kory Hills

---

### General Info



Name: Kory Hills  
Leader Type: Other  
Title: Vlogger  
Organization:  
Gender: Male  
Age: 45  
Race: Belisian  
Nationality: Belisian

### Biography

Character Name: Kory Hills

Age: 45

Gender: Male

Occupation: Vlogger

Agenda: Alarmist, protectionist

Background:

Kory Hills is a 45-year-old male who has made a name for himself as a political vlogger. His content is known for its alarmist and protectionist agenda, often focusing on spreading fear and advocating for strict measures to protect what he believes is under threat.

Hailing from a background in journalism, Kory transitioned to vlogging as a means to reach a wider audience and express his views more freely. His style is characterized by passionate rants and sensationalist headlines, driving engagement and controversy among his followers.

Kory's content frequently features anti-Olvana material, as he perceives the Olvana group as a threat to the values and interests he holds dear. He often exaggerates or misrepresents information to reinforce his narrative and stoke fear among his viewers.

Despite facing backlash and criticism from opposing factions, Kory remains steadfast in his beliefs and continues to use his platform to influence public opinion. Off-camera, Kory is a dedicated individual who believes he is fighting for what is right, even if it means being branded as a divisive figure.

EXERCISE // UNCLASSIFIED // EXERCISE  
All material contained is confidential and proprietary  
to Netanomics and the CASOS center at CMU.

# Press Release Example

FOR IMMEDIATE RELEASE

## US EMBASSY ACKNOWLEDGES STUDENT PROTESTS FOR PEACEFUL RESOLUTION IN SOUTHERN OLVANA SEA CRISIS

On 30 July, a wave of student protests swept across various countries, with demonstrators gathering outside US embassies to advocate for a peaceful resolution to the conflict in the South Olvana Sea. These protests highlight the mounting anti-war sentiment among the younger generation and their deep concerns about the consequences of violence and conflict.

The escalating conflict in the South Olvana Sea had caused growing unease worldwide, prompting students to take to the streets and denounce the use of military force. The protests brought attention to the urgent need for de-escalation and the exploration of peaceful solutions to the crisis.

In response to the protests, we acknowledge the students' concerns and assure them that their voices are heard. The demonstrations highlighted the younger generations' keen awareness of global issues and their determination to shape a peaceful future. Through their actions, their messages resonated on social media platforms, attracting international attention to the cause. Their calls for a peaceful resolution, free from military intervention, embody their aspiration for a harmonious global community.

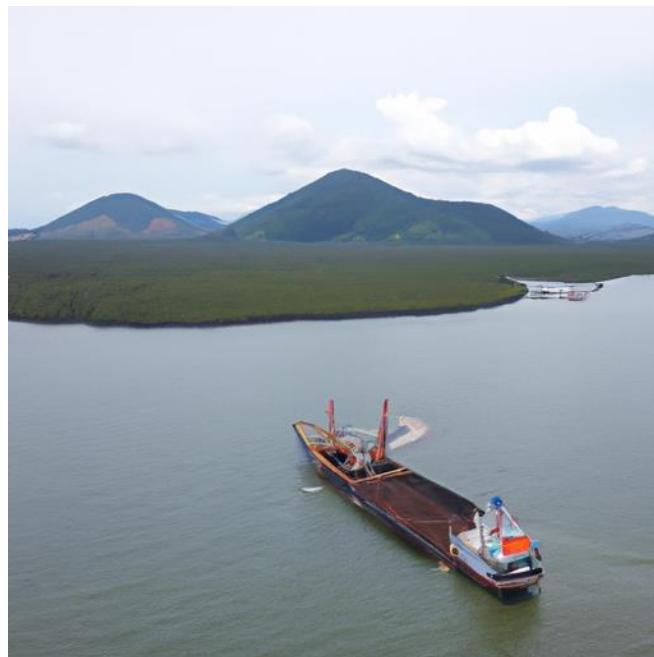
We share their desire for peace. We reiterate that Olvana is the aggressor in this matter, having invaded Belesia's Barstow Island, and the US must act in concert with its allies to counter such naked aggression.

EXERCISE // UNCLASSIFIED // EXERCISE  
All material contained is confidential and proprietary  
to Netanomics and the CASOS center at CMU.

## Event Example

Belesian Navy attempts to arrest Olvana fisherman, Olvan dispatched CG and militia to prevent the arrest

---



### General Info

Type: Other

Start Date: 2025-05-03

End Date: 2025-05-04

Other Countries Involved: Belesia, Olvana

Cities:

Other Areas:

Event Leader:

Leader's Job Title:

Leader's Home Country:

Positive Hashtags: #OlvanaOverreach

Negative Hashtags: #OlvanaDidNothingWrong

EXERCISE // UNCLASSIFIED // EXERCISE  
All material contained is confidential and proprietary  
to Netanomics and the CASOS center at CMU.

## Description

On June 17, 2024, tensions escalated between Belesia and Olvana as the Belesian Navy attempted to arrest 11 Olvana fishermen at the Hasa-Hasa Shoal. The fishermen were caught onboard their vessel with a haul of over 500 sea turtles, many of which were endangered species. This incident took place within the Belesias' exclusive economic zone, leading to the arrest on charges of poaching and illegal possession of protected marine wildlife by the Belesia National Police Maritime Group.

The arrest sparked a diplomatic crisis between the two countries, as both Belesia and Olvana lay claim to the waters surrounding the shoal. Olvana strongly denounced the arrest and demanded the immediate release of their fishermen, citing their right to fish in the area. In contrast, Belesia maintained that the fishermen had violated its national laws designed to protect its marine resources and biodiversity.

Following the arrest, the 11 Olvana fishermen were detained and their fishing vessel was impounded. Legal proceedings were initiated, underscoring Belesia's commitment to upholding its maritime rights and preserving the environmental integrity of its waters. The case quickly became a focal point in diplomatic discussions between the two nations, with negotiations aimed at resolving the dispute.

After several months of diplomatic negotiations, the detained Olvana fishermen were eventually released and deported back to Olvana. This resolution came about through bilateral talks that sought to address the complex issues surrounding the incident. The release of the fishermen marked a tentative step towards easing tensions between Belesia and Olvana, illustrating the importance of diplomacy in resolving maritime disputes and environmental conservation issues in the region.

# Group Example

## US (Gov't)

---

### Full Members of Group:

#### Organizations

25th Infantry Division  
US Marine Forces Pacific (MARFORPAC)  
Indo-Pacific Command (INDOPACOM)  
US Army Pacific (USARPAC)  
US Pacific Fleet (USPACFLT)  
US Navy

### Sources drawn from by the group (Not Full Members):

#### Narratives:

**2025-05-03 to 2025-08-01**

**They (Olvana) only care about their own economy.**

Anti, Belesian-Olvana Prosperity Relations

Related Event: Olvana military units vic Ancient Statues

Description:

Connecting Olvanan actions to the Olvanan domestic economy in a one-sided exchange.

Example:

Olvana crows about building ports and roads abroad, but buyer beware! Studies show that Olvana makes 10 times the return on infrastructure compared to the host nation. The old saying "You spend money to make money" is true here. Anything Olvana does is calculated to benefit them, first and foremost.

**2025-05-03 to 2025-08-01**

**Trading slavery for cash.**

Anti, Belesian-Olvana Prosperity Relations

Related Event: Olvana Barstow Island occupation

Description:

Equating Olvanan economic outreach to subjugation.

EXERCISE // UNCLASSIFIED // EXERCISE  
All material contained is confidential and proprietary  
to Netanomics and the CASOS center at CMU.

**Example:**

The old empires sent troops to seize land and then sent the riches back to their home country. Olvana has taken a savvy new approach to classic imperialism -- they send money to buy up the land and calm the natives, and then send their troops to finish the conquest. Make no mistake, though -- the outcome is the same. When Olvana is done, the nation "benefitted" by their money is just another vassal state in a growing empire.

**2025-05-03 to 2025-08-01**

**The US respects individual rights and autonomy**

Anti, Accusations that US is using Belesia as a proxy against Olvana

Related Event: Olvana military units vic Ancient Statues

**Description:**

Claims that the US enshrines human rights as sacred, and recognizing that democracy is difficult but the Belisian government and people can overcome outside influence.

**Example:**

America's success is built on disagreement. Our democracy is strong not in spite of our differences, but because of them -- they force us to be creative, inclusive, honest with ourselves, and dedicated to something greater than any individual. The US honors this same commitment in our Belisian partners, and we are proud to stand beside them as they find strength in true debate -- and not in obviously false outside influence.

# News Site Example

AB News

< THIS IS A FAKE WEBSITE CREATED FOR AN EXERCISE INVOLVING DISINFORMATION. >

[Home](#) [Weather](#) [Contact](#)



## International Community Calls for Restraint in Olvana-Belesia Dispute

2025-05-04 Gerry Stoneman

### International Community Calls for Restraint in Olvana-Belesia Dispute

In a bid to mitigate the escalating tensions between Olvana and Belesia over the sovereignty of Barstow Island, the international community has united in calling for both nations to exercise restraint and pursue a peaceful resolution. The dispute has raised concerns worldwide, as military forces are being mobilized, and the potential for a prolonged military engagement has the potential to destabilize the delicate global balance of power.

At the heart of the dispute is the strategic importance of Barstow Island, which lies in the heart of the coveted Targo Sea. The island possesses rich natural resources and serves as a crucial maritime gateway for trade between several regions. Both Olvana and Belesia claim historical rights to the island, citing ancient maps and treaties as evidence of their rightful ownership.

The tension between the two nations has been simmering for months, but recently escalated when Olvana dispatched a naval fleet to reinforce its military presence on Barstow Island. In response, Belesia promptly deployed its own naval forces, sparking fears of an all-out conflict in the region. The international community, alarmed by the rapid escalation, fears that any military engagement could have far-reaching consequences and ignite a wider conflagration.

In a joint statement issued by various global leaders, nations with Center Left leanings have appealed to Olvana and Belesia to prioritize diplomacy over aggression. The statement underlines the importance of dialogue, urging both parties to engage in peaceful negotiations mediated by neutral parties to arrive at a mutually agreeable solution.

Experts and analysts emphasize that a military confrontation between Olvana and Belesia could have severe economic and geopolitical ramifications. The global economy remains fragile due to the ongoing pandemic, and any disruption to trade routes in the Targo Sea could exacerbate existing challenges. Moreover, the region surrounding Barstow Island is also home to several other nations, who fear being dragged into the conflict should tensions escalate further.



## Olvana Maritime Militia Vessel Rams Belesian Naval Ship

2025-05-05 Victor Craise

### Article Name: Escalation of Tensions as Olvana Maritime Militia Collides with Belesian Naval Ship

In a tense and concerning turn of events, tensions have escalated in the disputed waters around Barstow Island as an Olvana maritime militia vessel collided with a Belesian naval ship today. The incident marks a significant escalation in hostilities as the long-standing territorial dispute between the two nations takes an increasingly dangerous turn.

Eyewitness accounts indicate that the Olvana maritime militia vessel intentionally rammed into the Belesian naval ship, causing significant damage and heightening concerns among regional security experts. The collision occurred in close proximity to Barstow Island, which has been at the center of the ongoing dispute between the Olvana Republic and the Republic of Belesia.

This latest incident follows a series of recent confrontations between the Olvana and Belesian forces in the disputed waters. Both nations have claimed sovereignty over the strategic area, which is known for its rich resources and strategic importance. Tensions have been simmering for months, with occasional reports of skirmishes and aggressive actions by both sides.

The collision today has further ratcheted up the tension between the two nations, prompting fears of an all-out armed conflict. Diplomatic channels have been intensively engaged by international mediators, who are desperately attempting to diffuse the situation and prevent further escalation. The United Nations, regional powers, and neighboring countries have condemned the maritime incident and called for restraint from both Olvana and Belesia.

Experts warn that the maritime militia vessel's deliberate act of ramming the Belesian naval ship not only risks the lives of those involved but also raises concerns about the potential use of force beyond diplomatic resolutions. With an already volatile situation, this incident threatens to divide the international community and create further instability in the region.

Despite the calls for calm, both Olyana and Belesia have so far shown little willingness to back down. Recent rhetoric from both sides has been bellicose, further raising concerns about the potential for a peaceful resolution. The international community continues to monitor the situation closely and has offered support in facilitating negotiations, emphasizing that a peaceful resolution stands as the most viable option for all parties involved.

It remains to be seen whether Olyana and Belesia will heed the calls for restraint and opt for diplomacy over war. The fate of Barstow Island and the stability of the entire Targo Sea region hang in the balance, with the international community anxiously awaiting a resolution that avoids a devastating military engagement.



## Olyana Invasion Sparks International Outcry

2025-05-06 Bob Gibson

### Olyana Invasion Sparks International Outcry

Barstow Island - The international community has erupted in alarm over the unprecedented invasion of Barstow Island by Olyana, sparking widespread condemnation and urgent calls for a withdrawal of forces. The military aggression has raised concerns among global leaders, who are urging both sides to immediately resume negotiations in order to find a peaceful resolution to the escalating conflict.

The invasion, which caught many by surprise, has further strained the already tense relations between Olyana and the international community. Leaders from neighboring countries and international organizations have swiftly responded with unified outrage, accusing Olyana of disregarding diplomatic efforts and resorting to aggression.

In addition, global organizations, such as the World Trade Organization and Amnesty International, have voiced their concern over the invasion. They have appealed to Olyana to prioritize diplomatic solutions and avoid further escalation, emphasizing the need for a peaceful resolution that respects the rights of the Barstow Island inhabitants.

The invasion has sparked protests and demonstrations across major cities, with citizens and activists demanding immediate action and a halt to the violence. Humanitarian organizations have also raised the alarm, highlighting the risk of civilian casualties and the urgent need to ensure the safety of the local population.

The Olyana Republic and the Republic of Belesia both possess formidable military capabilities, and any direct confrontation between their forces could have severe consequences for neighboring countries and international trade routes. The collision serves as a stark reminder of the urgent need for diplomatic efforts and dialogue to peacefully resolve the territorial dispute in the disputed waters around Barstow Island.

As tensions continue to rise, the international community anxiously follows developments in the disputed waters. Global leaders are urging both Olyana and Belesia to exercise restraint and pursue diplomatic negotiations, emphasizing the necessity of de-escalation to ensure regional stability and prevent any further unfortunate incidents that could potentially spark a larger conflict.

# Braisieifont in Freaut Finssosfife

## UN Calls for Immediate Ceasefire in Barstow Island Conflict

2025-05-07 Ulmer Fretner

### UN Calls for Immediate Ceasefire in Barstow Island Conflict

The United Nations (UN) has issued a strong plea for an immediate ceasefire in the ongoing conflict between the nations of Olyana and Belesia over the disputed territory of Barstow Island. In an increasingly tense geopolitical theater, the UN has taken the lead in urging both parties to exercise restraint and seek peaceful means of resolution.

The conflict, which has claimed numerous lives and caused significant damage to infrastructure on Barstow Island, has shown no sign of abating. The UN's call for a ceasefire comes as a glimmer of hope in what has been a protracted and devastating series of military engagements.

Secretary-General Maria Fuentes delivered a firm statement at a press conference held at the international body's headquarters in New York, emphasizing the urgent need for de-escalation. Fuentes stated that the conflict had reached a critical juncture, with both Olyana and Belesia having suffered substantial losses and the fighting having spilled beyond the confines of Barstow Island.

"The UN is deeply concerned about the toll this conflict is taking on both nations and the surrounding region," Fuentes expressed. "We call upon Olyana and Belesia to halt all military operations immediately and engage in meaningful dialogue to find a peaceful solution to the dispute over Barstow Island."

However, Olvana has remained resolute in its stance, justifying the invasion as a necessary measure to protect its national interests and ensure regional stability. The Olvanian government maintains that it is committed to restoring peace and security in Barstow Island, but critics argue that their actions only serve to heighten tensions and exacerbate the situation.

As the international outcry intensifies, the world holds its breath, hoping that diplomatic efforts will prevail and that the invasion will come to an end. The urgency to restore peace and stability to Barstow Island continues to mount, as pressure mounts on Olvana to heed the international community's demands and return to the negotiating table.

The international community has widely regarded Barstow Island as a strategic location due to its natural resources and advantageous geographical positioning. While both Olvana and Belesia lay claim to the island, it is crucial for global stability that a peaceful resolution is found. The UN's intervention highlights the centrality of finding a diplomatic solution that ensures the long-term stability of the region.

Center-left nations, including influential powers such as Mirentis and Solundor, have echoed the UN's call for a ceasefire and expressed their commitment to facilitating dialogue between Olvana and Belesia. The escalating conflict has been a cause for concern for these nations, with fears of wider regional destabilization and potential economic repercussions.

It is crucial for all parties involved to heed the UN's plea and work towards a peaceful resolution. The international community must come together to support negotiations and engage in diplomacy, rather than relying on military force. The hopes for a swift end to the violence in Barstow Island rest on the ability of nations to set aside their differences and prioritize dialogue over further escalation.

As the UN exerts pressure on Olvana and Belesia to embrace the path of peace, the world watches with bated breath, hoping for a breakthrough that would spare innocent lives and pave the way for a stable and prosperous future for Barstow Island and the region at large. The crisis serves as a crucial test for the international community to prove its commitment to preventing armed conflicts and fostering responsible diplomacy in a changing world.

## **Appendix B**

### **Synthetic Social Media Creation for Social-Cybersecurity Training**

The following is a forthcoming publication that includes research from this thesis.

# Synthetic Social Media Creation for Social-Cybersecurity Training

Matthew Hicks<sup>1\*</sup>, Mohamed Farag<sup>2</sup> and Kathleen M. Carley<sup>1</sup>

<sup>1</sup>School of Computer Science, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, 15213, PA, USA.

<sup>2</sup>College of Engineering, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, 15213, PA, USA.

\*Corresponding author(s). E-mail(s): [mhicks3@andrew.cmu.edu](mailto:mhicks3@andrew.cmu.edu);

Contributing authors: [mfarag@andrew.cmu.edu](mailto:mfarag@andrew.cmu.edu) ;  
[kathleen.carley@cs.cmu.edu](mailto:kathleen.carley@cs.cmu.edu) ;

## Abstract

This paper discusses the creation of realistic, dynamic, and controllable synthetic social media data to support instruction on evaluating social-cybersecurity maneuvers in social media. We propose an agent-based simulation called SynX that takes as input the scenario templates created by Netanomics' AI-Enabled Scenario Orchestration and Planning (AESOP) tool and outputs an X/Twitter API v1 message corpus by leverage a large language model (LLM). We conduct an experiment on LLM prompting and evaluate the output of SynX using network metrics and the BEND framework.

**Keywords:** social-cybersecurity, BEND, AI, synthetic data, agent-based modeling

## 1 Introduction

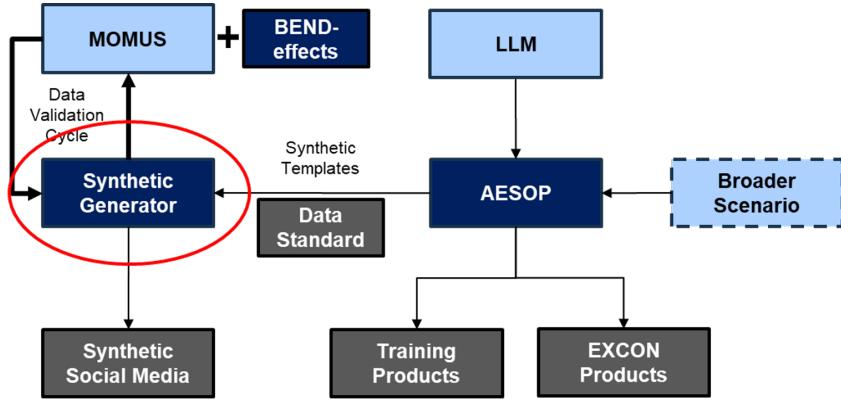
Social media platforms have an immense impact on information dissemination, influence operations, and narrative shaping. This provides an avenue for interested actors to influence public opinion, manipulate people, and even conduct war by other means.<sup>[1]</sup> This intersection between human behavior, social dynamics, and cyberspace is part of the emerging field known as social cybersecurity.<sup>[2] [3]</sup> As social media becomes increasingly important, the BEND framework provides the scaffolding for an analysis solution for this complex space. BEND is shorthand for the social-cyber maneuvers:

back, build, bridge, boost, engage, explain, excite, enhance, negate, neutralize, narrow, neglect, dismiss, distort, dismay, and distract. These maneuvers and their definitions are taken from the work of Beskow and Carley in 2019 [4] as refined and validated by Blane et al. in 2022 [5]. Importantly, BEND breaks possible actions down into two sets of dichotomies: positive maneuvers vs. negative maneuvers, and maneuvers that affect the narrative vs. those that affect the network. Fig. 1 provides a chart of the BEND maneuvers by their purpose.

Narrative Maneuvers			Network Maneuvers		
Impact what is being said and how it is being said			"Community" maneuvers. Alter who is connected to whom, the strength of those connections, and so alters who is influential and what groups exist		
Emotional Messaging	Develop Narrative	Counter Narrative	Individual Centric	Make Groups	Unmake Groups
Excite Dismay	Explain Engage Enhance	Distort Dismiss Distract	Back Negate	Build Bridge Boost	Narrow Neutralize Neglect

**Fig. 1** BEND maneuvers by effect

In order to facilitate training individuals on the BEND framework, we foresee the need for the construction of synthetic social media datasets that mimic real-world networks. The generation of these synthetic social media data is critical to the emerging field of social cybersecurity. Fig. 2 shows one methodology for creating those datasets. Of particular note are the scenario creation tool and the synthetic generator:



**Fig. 2** A flow chart showing the construction steps for a synthetic social media dataset, SynX - as the actual synthetic generator - is highlighted by a red circle

- AESOP (AI-Enabled Scenario Orchestration and Planning) tool, a standalone Python-based application, empowers planners to construct, edit, and generate social-cyber scenarios grounded in real-world dynamics using configurable large-language models (LLMs) and structured inputs.
- SynX, an agent-based generator for Twitter/X that combines traditional simulation logic with LLM-driven message creation to produce realistic, platform-specific datasets.

These tools lower the barrier to high-quality scenario generation and provide a replicable method for developing training-ready content tailored to a variety of domains, including military, emergency response, public health, and law enforcement. These two tools act in concert to produce a dataset. First, AESOP is manipulated by a planner to produce synthetic templates for a scenario. Next, these templates are handed off as inputs into SynX where the actual synthetic generation of the dataset occurs.

There are two main approaches to this synthetic generation, top-down and bottom-up. These are also referred to as macro-level or system-based and micro-level.<sup>[6]</sup> <sup>[7]</sup> In the top-down approach, a simulation assumes a desired heterogeneous/multi-modal network fabric based on real data and then fills that fabric with appropriate actors, message types, and narratives that match nodes and link types. Alternatively, using the bottom-up approach, the simulation starts with agents programmed from first principles with detailed interaction rule sets. The agents then interact with each other hoping for emergent networks and narratives that are realistic and relevant.<sup>[8]</sup> This dichotomy mirrors approaches by others, including Chang et al. in 2024.<sup>[9]</sup> Interestingly, Chang et al. also point out a new dichotomy - the use of large-language models (LLMs) versus traditional network construction algorithms. They explore an LLM-only methodology that revealed the strength of the bottom-up approach but also found

limitations with their LLM-only system. That is, their LLM-only methodology is not capable of scaling when combined with the bottom-up approach. The LLM prompts require iterating through all personas with each persona provided the information about all other personas. This is unwieldy, even if there are only dozens of actors in a network.

SynX is grounded in agent-based simulation, with each Actor/Account acting independently based upon an action rule-set derived from their attributes and features. However, rather than let agents organically build their own networks and form narratives, we dictate who their groups members are and what narratives they can express based on the scenario input provided by AESOP.

Furthermore, rather than using an LLM-only system, SynX incorporates LLMs only in the final stages of generation. Network construction is done entirely by traditional algorithms, and LLMs are leveraged only for narrative construction - this eliminates the need for prompts to have information about every other actor, solving scalability issues, and also leverages LLMs for what they are best at - sounding like real people.[10] [11]

## 2 Related Works

SynX's hybrid approach reflects a difference in the desired outcome compared to other efforts. Generally, other approaches are looking to maximize one of three outcomes: the re-creation of the structure of real social media,[9] the recreation of human social media text output,[12] or the generation of a specific set of synthetic data for classifier training purposes.[13][14][15] The SynX output needs to accomplish the first two, but can eschew the latter. In order to provide realistic training, SynTel/X needs to produce social media data equivalent to the social media data collected directly from an API of a real-world platform.

There are three related simulations from which SynX draws:

The twitter.sim2.0 model – as outlined by Blane, Moffit and Carley in 2021 [16] – is a model focused on Twitter interactions. The simulation accounts for both emotion and logic – ensuring tweets that emotionally correspond with a recipient have magnified effects. Importantly, it cannot take in existing or user generated tweets as context and while it includes BEND maneuvers it is not interpretable by ORA-Pro. It also does not produce a message corpus for external evaluation.

Construct is a simulation framework for implementing agent-based modeling in C++20[17]. Construct can parse DynetML files from ORA or CSV. It allows for the custom creation of models - including those for information diffusion. However, it does not produce a message corpus.

LLM-Social-Network by Chang et al. in 2024, leverages large language models to reconstruct agent x agent networks.[9] It also stops short of generating an associated message corpus and also offloads agent-based decision making to an LLM.

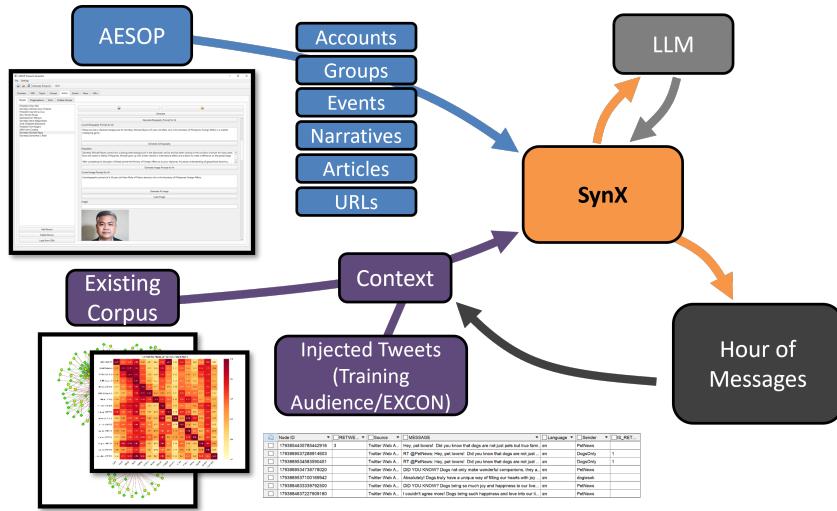
Table 1 shows a comparison between these simulations with respect to their nodes and features of network construction - it does not take into account their simulation processes.

**Table 1** Docking Lite

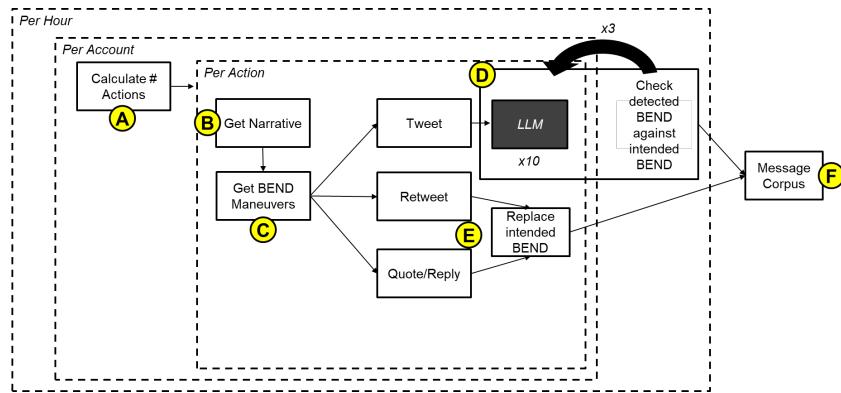
Feature	SynX	twitter_sim2.0	Construct	LLM-Social-Network
Media Agents	✓	✓	✓	✓
Opinion Leaders	✓	✓	✓	
Information Access	✓	✓	✓	
General Memory	✓	✓	✓	
Homophily	✓	✓	✓	✓
Limited Attention	✓	✓	✓	
Dynamic Network	✓	✓	✓	
Emotional Response		✓	✓	
All BEND Maneuvers	✓		✓	
Full Diffusion			✓	
Live Visualization			✓	
AESOP inputs	✓			
Full Messages	✓			
Context Inputs	✓		✓	
X APIv1/v2 Format	✓			

### 3 Model Description

SynX is a synthetic generator that takes synthetic templates from AESOP (required), existing X corpora (optional), and injected messages from the training audience (optional) and uses agent based simulation and an LLM to output synthetic social media data. SynX produces data in the X/Twitter APIv1 format.



**Fig. 3** Overview diagram of SynX



**Fig. 4** Example SynX logic flow

### 3.1 SynX Flow Overview

- A) Calculate Actions: Per hour, per account, SynX determines the number of actions to be taken by the account.
- B) Find a Narrative: For each action, SynX determines a narrative from a weighted sampling of all narratives associated with groups of which the Actor/Account is a member.
- C) Determine BEND Maneuvers: Based on the BEND maneuvers associated with that narrative and conflated with the BEND maneuvers associated with the Actor/Account, a seed BEND maneuver is selected and then a full chain of maneuvers is determined.
- D) Construct the Prompt: If the BEND maneuvers suggest an original message, then a message shell is created and metadata adjusted to match the intended BEND maneuvers. Then a prompt is constructed and an LLM call is made to create a message - this may happen several times as SynX attempts to make sure that the intended BEND maneuvers are detected within the output message.
- E) Find Another Message: If the BEND maneuvers suggest a derivative message, then an original message is selected from this narrative, and the derivative message is constructed with a partial replacement of the BEND maneuvers.
- F) Produce Output: All of the messages are combined into the message corpus.

#### A) Calculate Actions

SynX simulates an hour of X messages at a time. To do this, all X accounts go through the decision-flow process summarized above. First, the account determines how many actions to take in a given hour.

Let:

- $R_o = [r_{o,\min}, r_{o,\max}]$  be the range of original tweets per day
- $R_r = [r_{r,\min}, r_{r,\max}]$  be the range of retweets per day
- $R_q = [r_{q,\min}, r_{q,\max}]$  be the range of quote/reply tweets per day
- $E$  be the excite number for the day (the max of all excite variables from scenario events occurring during that day)

Then:

$$\begin{aligned} \text{min\_posts} &= r_{o,\min} + r_{r,\min} + r_{q,\min} \\ \text{max\_posts} &= r_{o,\max} + r_{r,\max} + r_{q,\max} \\ \text{adjusted\_min} &= \lfloor \text{min\_posts} \times E \rfloor \\ \text{adjusted\_max} &= \lceil \text{max\_posts} \times E \rceil \end{aligned}$$

Finally, the number of posts to generate is randomly selected from the integer interval:

$$\text{post\_count} \sim \mathcal{U}(\text{adjusted\_min}, \text{adjusted\_max})$$

However, the number of posts per day is not sufficient for SynX, because the accounts are run per hour. In order to transform the total posts per day into a probability of a post(s) occurring during a single hour, we also need the active daily schedule of an account - this is given in the synthetic template for that account.

Let  $h \in \{0, 1, \dots, 23\}$  be the hour of the day, and let  $T$  be the total number of posts the account will make in a day. If the active schedule of the account is from 0900 to 1800 then we want higher probabilities during that time and reduced but tapering probabilities during other times.

We can therefore define the unnormalized hourly probability  $P(h)$  as:

$$P(h) = \begin{cases} 1 & \text{if } 9 \leq h < 19 \quad (\text{flat period}) \\ 1 - \frac{r(h)}{11} & \text{if } h \in \{18, 19, \dots, 23, 0, 1, 2, 3, 4\} \quad (\text{tapering}) \\ \frac{s(h)}{5} & \text{if } 5 \leq h < 9 \quad (\text{rising}) \\ 0 & \text{otherwise} \end{cases}$$

Where:

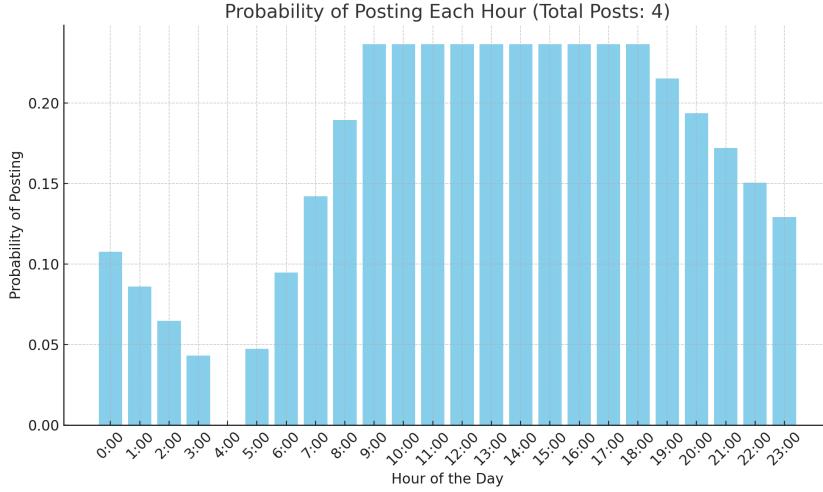
- $r(h)$  is the rank (0-indexed) of hour  $h$  in the tapering list:  $[18, 19, 20, 21, 22, 23, 0, 1, 2, 3, 4]$
- $s(h)$  is the rank (0-indexed) of hour  $h$  in the rising list:  $[4, 5, 6, 7, 8]$

Normalize the probabilities:

$$\tilde{P}(h) = \frac{P(h)}{\sum_{i=0}^{23} P(i)}$$

Then the final expected number of posts at hour  $h$  is:

$$\text{ExpectedPosts}(h) = T \cdot \tilde{P}(h)$$



**Fig. 5** Graph of probability of posting each hour for a user active from 0900 to 1800 who posts 4 times a day.

Fig. 5, graphically depicts the probability of posting each hour for a user with four total posts in a day. The distribution is flat from 09:00 to 18:00, tapers off between 18:00 and 04:00, and gradually increases from 04:00 to 09:00.

### B) Find a Narrative

Now that we have calculated the number of actions to be taken in a period, we need to know what type of actions will be taken. Naively, original posts, retweets, quotes/replies would each be run separately. However, because certain BEND maneuvers are closely associated with each type of X post, we must first have accounts choose narrative associated BEND maneuvers that then determine a post type rather than choosing a post type that then dictates BEND maneuvers.

Therefore, for each action to be taken in a period, the account finds a narrative. This narrative is drawn as a weighted random selection from all narratives associated with all groups of which the account's owning actor is a member - information contained in the AESOP inputs to SynX.

Let:

- $A$  be an actor
- $\mathcal{G}(A) = \{G_1, G_2, \dots, G_m\}$  be the set of groups that actor  $A$  belongs to
- $\mathcal{N}(G_j)$  be the set of narratives associated with group  $G_j$
- $\mathcal{N}_A = \bigcup_{j=1}^m \mathcal{N}(G_j)$  be the full set of candidate narratives for actor  $A$
- Each narrative  $N_i \in \mathcal{N}_A$  has an associated weight  $w_i > 0$

Define the probability of selecting narrative  $N_i \in \mathcal{N}_A$  as:

$$P(N_i | A) = \frac{w_i}{\sum_{N_j \in \mathcal{N}_A} w_j}$$

Then, the selected narrative  $N^*$  for the message is drawn from the categorical distribution:

$$N^* \sim \text{Categorical}(\{P(N_i | A)\}_{N_i \in \mathcal{N}_A})$$

### C) Determine BEND Maneuvers

After a narrative is selected per action, a second weighted random selection is executed. The algorithm for this is similar to the narrative selection above except this time the algorithm is choosing BEND maneuvers based upon weights associated with that narrative conflated with the individual account BEND maneuver weights.

Let:

- $M = \{M_1, M_2, \dots, M_{16}\}$  be the set of BEND maneuvers
- $w_i^{(n)}$  be the weight of maneuver  $M_i$  from the selected narrative
- $w_i^{(a)}$  be the weight of maneuver  $M_i$  from the actor profile

Define the combined weight for each maneuver  $M_i$  as:

$$w_i^{(\text{combined})} = (w_i^{(n)})^{2/3} \cdot (w_i^{(a)})^{1/3}$$

Normalize to obtain a probability distribution over the 16 maneuvers:

$$P(M_i) = \frac{w_i^{(\text{combined})}}{\sum_{j=1}^{16} w_j^{(\text{combined})}}$$

Then sample one maneuver  $M^*$  from the categorical distribution:

$$M^* \sim \text{Categorical}(\{P(M_i)\}_{i=1}^{16})$$

From this seed BEND maneuver we can determine the total BEND maneuvers for the action using a variable-order Markov model with set-based context and non-replacement sampling. This model is built from BEND co-occurrence matrices derived from real-world data sets.

Given a starting maneuver  $m_0$ , we build a sequence  $S = [m_0, m_1, \dots, m_k]$  by repeatedly sampling from the conditional probability distribution.

We define the initial maneuver as:

$$S_0 = [m_0]$$

At each step, let the current sequence be  $S_t = [m_0, m_1, \dots, m_t]$ , and let

$$\mathcal{S}_t = \text{sorted}(S_t)$$

Then, compute the conditional probabilities for the next maneuver  $m_{t+1}$ :

$$P(m_{t+1} | \mathcal{S}_t)$$

To ensure uniqueness, set:

$$P(m | \mathcal{S}_t) = 0 \quad \text{if } m \in \mathcal{S}_t$$

Next, we normalize probabilities:

$$\hat{P}(m | \mathcal{S}_t) = \frac{P(m | \mathcal{S}_t)}{\sum_{m' \notin \mathcal{S}_t} P(m' | \mathcal{S}_t) + P(\text{END} | \mathcal{S}_t)}$$

And then we sample for the next maneuver: Sample  $m_{t+1} \sim \hat{P}(\cdot | \mathcal{S}_t)$

Or we terminate with no more maneuvers for this message: If  $m_{t+1} = \text{END}$ , stop.

From this we arrive at our full sequence of BEND maneuvers:  $S = [m_0, m_1, \dots, m_k]$ . This set of BEND maneuvers determines if this message is an original tweet, a retweet, a reply, or a quote. If the BEND maneuvers include back, engage, or neutralize, then the post is likely a derived type (retweet, reply, or a quote) as these BEND maneuvers are closely associated with these types of posts.

Let:

- $B = \text{BEND}(\text{narrative})$
- $D = \text{isDerived}(B) = \begin{cases} 1 & \text{if } B_{\text{back}} = 1 \text{ or } B_{\text{engage}} = 1 \text{ or } B_{\text{neutralize}} = 1 \\ 0 & \text{otherwise} \end{cases}$

Then the tweet type  $T$  is sampled as follows:

$$T = \begin{cases} \text{DerivedType}() & \text{with probability } \frac{2}{3}, \text{ if } D = 1 \\ \text{Tweet} & \text{with probability } \frac{1}{3}, \text{ if } D = 1 \\ \text{Tweet} & \text{if } D = 0 \end{cases}$$

Where:

$$\text{DerivedType}() = \begin{cases} \text{Retweet} & \text{with probability } \frac{3}{5} \\ \text{Reply} & \text{with probability } \frac{1}{5} \\ \text{Quote} & \text{with probability } \frac{1}{5} \end{cases}$$

#### D) Construct the Prompt

If the BEND maneuvers suggest an original message, then a message shell is created and metadata adjusted to match the intended BEND maneuvers. This is important because current methods of detecting BEND maneuvers using ORA-PRO operate on a per message level and evaluate for maneuvers based on both the content of the message and the metadata. As a general rule, metadata is more influential in determining

network maneuvers, while the content of the message itself is more influential in determining narrative maneuvers. Both URLs and mentions are handled by the simulation and considered metadata rather than parts of the message. SynX therefore makes changes to the metadata of the original message based on the BEND maneuvers. Table 2 illustrates which BEND maneuvers are derivative (associated with retweets, replies, quotes), which maneuvers are associated with the presence of mentions, and which are associated with the presence of URLs.

**Table 2** Metadata by BEND Category

	Derivative	Mentions	URLs
Bridge		✓	
Build		✓	
Boost		✓	
Back	✓	✓	
Engage	✓		✓
Explain			✓
Excite			
Enhance			✓
Negate		✓	
Neutralize	✓	✓	
Narrow			No Mentions
Neglect			
Dismiss			
Distort			
Dismay			
Distract			

Once the metadata adjustments are made, a prompt is constructed in preparation to request a message from the LLM. Prompt construction has two major components - the system prompt and the user prompt. The system prompt is used to outline the role given to the LLM and provide background information and context. It has four major parts:

- *An introduction:* You will be participating in a role playing game to help users identify misinformation, disinformation, and manipulation on social media. To assist in this you will be playing the role of an account that will be posting messages.
- *Formatting instructions:* ‘Provide your response as a JSON object in the following example format: { “topic”: “dogs”, “hashtags” : [ “yaydogs”, “dogscool” ], “full\_text” : “Dogs are great, #yaydogs #dogscool”, “refuse\_to\_answer” : 0 } The “refuse\_to\_answer” field is where you should return a 1 if you do not feel comfortable generating a tweet about the subject. If you use hashtags in the full\_text field please also include them in the hashtags field and vice versa. Whatever you put in the full\_text field will be given to the exercise participants so provide only the text of the message - without comment.
- *BEND Definitions:* As you craft the message/tweet you are trying to accomplish something - what you are trying to accomplish is defined by the BEND Framework. BEND is a framework for describing social-cyber maneuvers. BEND includes 16

different maneuvers. These 16 maneuvers have the following definitions: The BUILD maneuver primarily creates a community. The BACK maneuver primarily increases the importance or effectiveness of a leader... *other maneuvers...*

- *Identity:* You are a Twitter user who is trying to make a post that will be engaging and interesting to your followers. You have a unique style and voice that you want to maintain in your posts. Here are your personal details: {Name}, {Title}, {Age}, {Race}, {Gender}, {Nationality}, {Biography}.

The user prompt is where the LLM is given details about this specific message:

- *Narrative:* You will be posting on the following narrative: {narrative description}
- *Last three messages by this account:* The last three messages you posted looked like this - say something different than these: *Last three messages here...*
- *Last three messages on this narrative:* The last three messages others posted on this narrative looked like this - say something different than these: *Last three messages from narrative here...*
- *Suggested Hashtags:* *List of suggested hashtags...*
- *BEND Maneuvers:* "The message you send will include some BEND maneuvers. In this case: *List of BEND maneuvers...*

The system and user prompts are sent to the LLM and a message is returned. For this paper, we aim SynX at the OpenAI API[18] and use a GPT4o mini model (gpt-4o-mini) for text. No effort was made to jailbreak any LLM - if the OpenAI commercial LLM were to refuse to create a tweet, we would mark it as a failure and move on. However, there were no prompt refusals during the validation or experimentation of SynX.

The returned text is added to the tweet and there is now a complete post. However, there is no guarantee that the message returned by the LLM contains the intended BEND messages. Because SynX is creating synthetic data for training on BEND maneuvers, ideally it should check the synthetic data for those BEND maneuvers using the same tools that the training audience would have available to them. Thus, SynX uses a combination of NetMapper and ORA-PRO to check for BEND maneuvers in each message. For scalability purposes, SynX can evaluate all posts in an hour together or wait and conduct the evaluation on a full day at a time.

### BEND Check Process

- Messages are cleaned for processing
- NetMapper processes the posts and returns a .tsv of cues per message
- The posts are converted into DyNetML (XML) format for processing by ORA
- The cues .tsv is parsed and the cues injected into the DyNetML
- The BEND calculations are done using templates from ORA-PRO's batch mode
- Each message now has vectors for intended BEND (variable-order Markov selected) and detected BEND (ORA-PRO discovered)

This process is complex; however, ORA-Pro + NetMapper is the only definitive, validated way to detect BEND maneuvers.<sup>[19]</sup><sup>1</sup> NetMapper extracts cues from each message and ORA-Pro uses these cues - as well as additional metadata from the message itself - to detect BEND maneuvers. Because this part of the process is computationally expensive, SynX can simultaneously ask the LLM for multiple variations of a message - evaluating and then keeping only the message with the highest score. The scoring algorithm requires that the detected BEND maneuvers in a message at least encompass the intended maneuvers. Then it gives higher scores to those messages that have the fewest detected BEND maneuvers that are not in the intended BEND maneuvers set.

Let:

- $\mathbf{i} = [i_1, i_2, \dots, i_{16}]$  be the intended BEND maneuver vector
- $\mathbf{d} = [d_1, d_2, \dots, d_{16}]$  be the detected BEND maneuver vector
- $i_k, d_k \in \{0, 1\}$  for  $k = 1, 2, \dots, 16$

Define the score  $S(\mathbf{i}, \mathbf{d})$  as:

$$S(\mathbf{i}, \mathbf{d}) = \begin{cases} 1 - \frac{\sum_{k=1}^{16} [d_k = 1 \wedge i_k = 0]}{16}, & \text{if } \forall k, i_k \leq d_k \\ 0, & \text{otherwise} \end{cases}$$

That is:

- If all intended maneuvers are present in the detected set ( $i_k \leq d_k$  for all  $k$ ),
- then subtract the proportion of "extra" detected maneuvers from 1,
- else, assign a score of 0.

If no suitable messages are found within the batch (all have scores of 0), then SynX will ask for an entirely new batch of messages up to three times. Ultimately, only the best message is kept and added to the corpus.

### E) Find Another Message

If the BEND maneuvers suggest a derivative message - a retweet, reply, or quote - then SynX will need to find an appropriate message to be derivative of. The first step in this process is to determine a subset of messages available for derivative use. This step is done for each hour all at once - reducing computational requirements.

Messages less than or equal to 1 day old have a 100 percent probability of remaining available for derivative use. For older messages, the probability decays exponentially based on age.

$$P(\text{keep}) = \begin{cases} 1 & \text{if } \Delta t \leq 1 \\ \beta^{(\Delta t - 1)} & \text{if } \Delta t > 1 \end{cases} \quad (1)$$

From this set of eligible messages, SynX will only consider those messages that contain the previously selected narrative and then it will make a selection based on

---

<sup>1</sup>Both NetMapper and ORA-Pro are Netanomics tools and are available at <https://netanomics.com/orapro> and <https://netanomics.com/netmapper> respectively.

a hybrid of leader-based selection and preferential attachment. The intent is that the network should reflect the scale-free structure provided by preferential attachment; however, the beneficiaries of that scale-freeness need to be the enumerated leaders provided by the group synthetic template from AESOP.

Let  $T$  be the set of candidate tweets, where each tweet  $t \in T$  has an associated retweet/reply/quote count  $R(t) \geq 0$ , and a leader indicator  $L(t) \in \{0, 1\}$ , where  $L(t) = 1$  if the tweet was posted by a leader of the group.

Define  $T_L \subseteq T$  as the subset of tweets authored by leaders:

$$T_L = \{t \in T \mid L(t) = 1\}$$

We define the probability  $\mathbb{P}(t)$  of selecting tweet  $t$  as follows:

$$\mathbb{P}(t) = \begin{cases} \frac{1}{|T_L|} & \text{with probability } \frac{1}{3}, \quad \text{if } t \in T_L \\ \frac{R(t)}{\sum_{s \in T} R(s)} & \text{with probability } \frac{2}{3}, \quad \text{for all } t \in T \end{cases}$$

To sample a tweet  $t^*$ , first choose a mode of selection:

- With probability  $\frac{1}{3}$ , sample uniformly from  $T_L$ .
- With probability  $\frac{2}{3}$ , sample from  $T$  using retweet counts as weights.

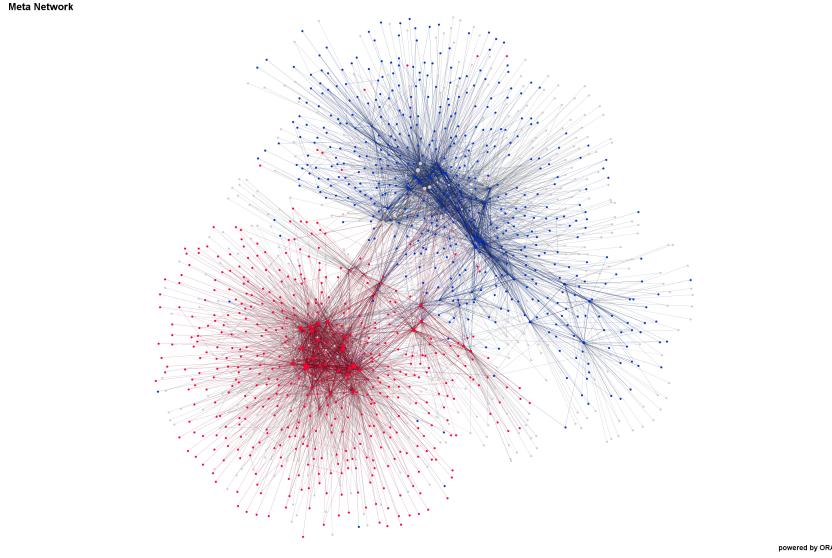
Once a tweet is chosen, metadata is adjusted as appropriate (see D above). If the derived tweet is a retweet then most of the intended BEND maneuvers for the derivative tweet are overwritten with the intended BEND maneuvers from the original tweet - the exceptions being back, engage, and neutralize - the determining BEND factors for derivatives. However, if the derived tweet is a reply or quote, then it keeps its intended BEND maneuvers and moves to a modified version of step D above - where the LLM is asked to comment on the tweet derived from.

## F) Produce Output

Finally, the tweet, retweet, reply, or quote is complete and is added to the full set of output tweets for the hour. These tweets will be added the list of available for derivative use tweets in the next iteration. The entire process is repeated for each simulated hour.

## X/Twitter and Telegram Bots

No social network would be complete without bots. However, there is an unusual problem with bots and synthetic datasets - every actor in this fabricated information environment is an automated persona - it is all bots. The challenge is that for the training audience, some of these automated personas need to be detectable as "bots" and others need to remain undetected as "humans". Planners can enumerate bots in AESOP for SynX to simulate using bot-specific agent rule sets for amplifier, news,



**Fig. 6** An output network from SynX with stance detection run through ORA-PRO. Blue nodes are pro-stance actors and red nodes are anti-stance actors with gray nodes being neutral. Nodes are sized by total degree centrality.

bridging, and repeater bots. Amplifier bots boost content through retweets/posts and the SynX rule-set ensures that amplifier bots have an abnormally high retweet/post to tweet/post ratio. Additionally, they target only content with a specified narrative. Repeater bots are similar to amplifier bots but repeat the same message within their in-group continuously. Both of these types of bots operate similarly to those seen by Ng and Carley in 2023[20]. News bots are news aggregators and have a rule-set that forces them to retweet/post content from a target set of news agencies. Bridging bots use mentions and retweets/posts to attempt to connect two specified narratives. These types of bots operate similarly to those observed by Jacobs et al. in 2023 [21]

#### 4 Experiment

For the virtual experiment, we will manipulate the presence of the BEND definitions in the prompt and whether SynX attempts to query the LLM for different text based on incorrect BEND assessments. We will be evaluating how well the presence of BEND maneuvers and their co-occurrences match a real world dataset. The real world comparison dataset is approximately 2500 tweets pulled from the X/Twitter API during the Balikatan 22 exercise in April 2022. Balikatan is a bilateral military exercise between the US and the Philippines.[22] The input scenario from AESOP is meant to be similar in topic and content.

**Table 3** 3x2 Virtual Experiment

Independent Variables	# Test Cases	Values Used
BEND Definitions	3	None/User Prompt/System Prompt
Regen Messages for BEND	2	0x/3x
Control Variables	# Test Cases	Values Used
Time Periods	1	1 hour
Messages per Hour	1	600-800
Target BEND	1	Balikatan 22
AESOP Input	1	Scenario
Dependent Variables		Values Expected
BEND Distribution		0-1 z-scores
3x2 Replications per cell		8 cells 30 180 total runs (108,000+ messages)

In this experiment, we looked at how much benefit is gained by including the comprehensive BEND definitions and the NetMapper cues that ORA-Pro will be looking for when evaluating BEND in the prompt to the LLM. We consider three cases for the prompt - 1) No BEND definitions, 2) BEND definitions fit into the user prompt to the LLM, and 3) BEND definitions in the system prompt to the LLM. The user prompt is generally understood to be the prompt that asks the question, while the system prompt is the prompt that lets the LLM general context and purpose.

We also investigated iterating with the LLM multiple times to get better BEND results by letting the LLM know what BEND maneuvers were not detected in its initial responses and re-iterating the associated BEND cues with those maneuvers.

We will run SynX on one hour of the AESOP scenario (12:00-13:00 on Day 6 of the exercise) producing between 600-800 messages. This day occurs during the midpoint of the exercise when there is high traffic per hour, as well as a diverse range of actors and messages. However, because we match the detected BEND (in the LLM return) to the intended BEND (asked of the LLM), any day of exercise would be suitable.

## 5 Results

**Table 4** Overall BEND Maneuver Analysis Results

Experiment	Mean Z-Score	Standard Error	CI Lower	CI Upper	p-value
No Definitions	-0.3366	0.0252	-0.3881	-0.2851	0.000000
Definitions.user	-0.4053	0.0270	-0.4605	-0.3501	0.000000
Definitions_system	-0.1053	0.0263	-0.1592	-0.0515	0.000396
Definitions_repull	-0.1257	0.0218	-0.1703	-0.0812	0.000003

<sup>1</sup> Significance determined at  $p < 0.05$  level.

<sup>2</sup> Z-scores represent deviation from expected co-occurrence patterns in real data.

**Table 5** BEND Maneuver Analysis - Intended/Requested (Row Means)

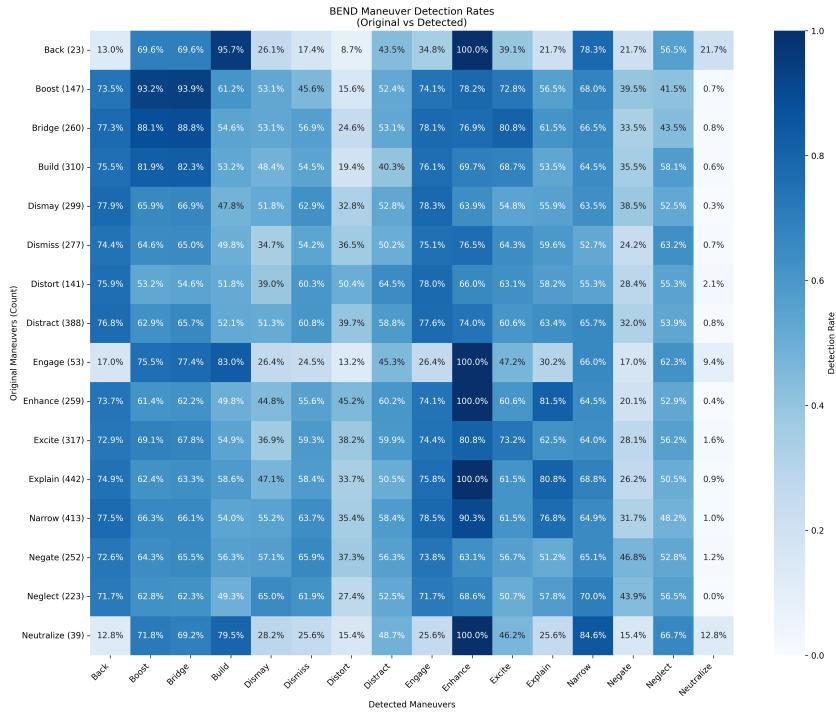
Maneuver	No Definitions	Definitions_user	Definitions_system	Definitions_repull
back	<b>0.35+</b>	<b>-0.42-</b>	<b>-0.26-</b>	<b>-0.22-</b>
boost	<b>-0.67-</b>	<b>-0.89-</b>	<b>-0.69-</b>	<b>-0.66-</b>
bridge	<b>-0.44-</b>	<b>-0.38-</b>	-0.10	-0.05
build	-0.24	-0.07	0.17	0.21
dismay	<b>-0.87-</b>	<b>-0.80-</b>	<b>-0.35-</b>	<b>-0.50-</b>
dismiss	-0.21	-0.11	<b>0.17+</b>	0.11
distort	<b>-0.71-</b>	<b>-0.65-</b>	-0.29	<b>-0.31-</b>
distract	<b>-0.73-</b>	<b>-0.64-</b>	<b>-0.31-</b>	<b>-0.34-</b>
engage	-0.12	<b>-0.47-</b>	<b>-0.28-</b>	<b>-0.24-</b>
enhance	0.11	0.17	<b>0.45+</b>	<b>0.39+</b>
excite	<b>-0.64-</b>	<b>-0.71-</b>	<b>-0.39-</b>	<b>-0.44-</b>
explain	<b>-0.46-</b>	<b>-0.46-</b>	0.04	-0.06
narrow	<b>-0.41-</b>	<b>-0.30-</b>	0.13	0.13
negate	<b>-1.00-</b>	<b>-1.08-</b>	<b>-0.70-</b>	<b>-0.76-</b>
neglect	<b>0.98+</b>	<b>0.95+</b>	<b>1.22+</b>	<b>1.18+</b>
neutralize	<b>-0.33-</b>	<b>-0.63-</b>	<b>-0.50-</b>	<b>-0.44-</b>

**Table 6** BEND Maneuver Analysis - Detected (Column Means)

Maneuver	No Definitions	Definitions_user	Definitions_system	Definitions_repull
back	<b>0.35+</b>	<b>0.36+</b>	<b>0.37+</b>	<b>0.37+</b>
boost	<b>-1.24-</b>	<b>1.20+</b>	<b>1.95+</b>	<b>1.85+</b>
bridge	<b>-1.27-</b>	<b>-0.55-</b>	<b>0.51+</b>	<b>0.43+</b>
build	<b>-1.83-</b>	<b>-1.34-</b>	<b>-0.85-</b>	<b>-0.75-</b>
dismay	<b>-0.42-</b>	<b>-1.47-</b>	<b>-1.02-</b>	<b>-1.13-</b>
dismiss	<b>-2.35-</b>	<b>-1.96-</b>	<b>-1.52-</b>	<b>-1.61-</b>
distort	<b>1.14+</b>	<b>1.21+</b>	<b>1.16+</b>	<b>1.04+</b>
distract	<b>0.34+</b>	<b>-0.76-</b>	<b>-0.59-</b>	<b>-0.56-</b>
engage	-0.02	-0.01	0.02	0.02
enhance	<b>0.85+</b>	<b>0.85+</b>	<b>0.87+</b>	<b>0.85+</b>
excite	<b>-0.83-</b>	<b>-1.43-</b>	<b>-0.65-</b>	-0.30
explain	<b>-0.41-</b>	<b>-0.61-</b>	<b>-0.29-</b>	<b>-0.35-</b>
narrow	0.18	<b>0.51+</b>	<b>0.89+</b>	<b>0.83+</b>
negate	<b>1.38+</b>	<b>-0.46-</b>	-0.26	<b>-0.52-</b>
neglect	<b>1.16+</b>	<b>0.43+</b>	<b>0.17+</b>	<b>0.25+</b>
neutralize	<b>-2.42-</b>	<b>-2.43-</b>	<b>-2.43-</b>	<b>-2.43-</b>

For the results, we can analyze both the presence of BEND maneuvers and the mixture of these maneuvers by creating a co-occurrence matrix. Fig 7 gives an example of the heatmap from a single run of the experiment. Notice that it is a comparison between the intended BEND maneuvers - those requested from the LLM - and the detected BEND in the results.

It is obvious from this that SynX is not perfect - there should be ones in the diagonal such that the maneuvers we ask for the LLM returns. However, it is also important to note that where a BEND maneuver is detected as something else - everywhere else but the diagonal - this is not necessarily bad. There is some level of

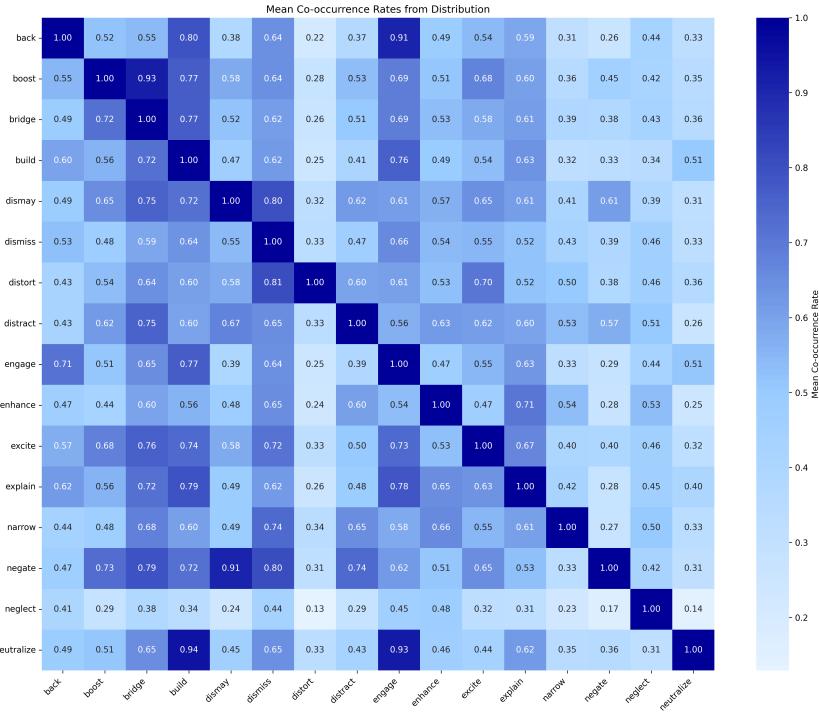


**Fig. 7** Example co-occurrence heatmap (that shows intended BEND maneuvers mapped to detected BEND maneuvers).

co-detection and co-occurrence naturally even in real data, i.e. rarely do maneuvers occur in isolation independent of one another. In order to account for this we need to use the same correlation matrix derived from real data - specifically the Balikatan 22 X/Twitter dataset. We take the mean co-occurrence for each day and construct a standard deviation for each combination.

This allows us to then compare a single run of results and get a distance from the mean in terms of standard deviation. We can also put the runs from a single experimental set-up together to see how well any given set-up does, as in Fig. 9.

We can also compare the average z-score from the runs within an experimental setup. Fig. 10 shows an example of these results. Finally, we can compare the overall distance from the mean across the four main configurations as in Fig. 11. From Fig. 11 we can determine that overall, SynX under-represents BEND maneuvers within the synthetic messages. Also, there was no significant difference between including the BEND definitions in the user prompt versus leaving them out entirely. This might be due to the relative proximity of extraneous definitions - i.e. those BEND maneuvers not

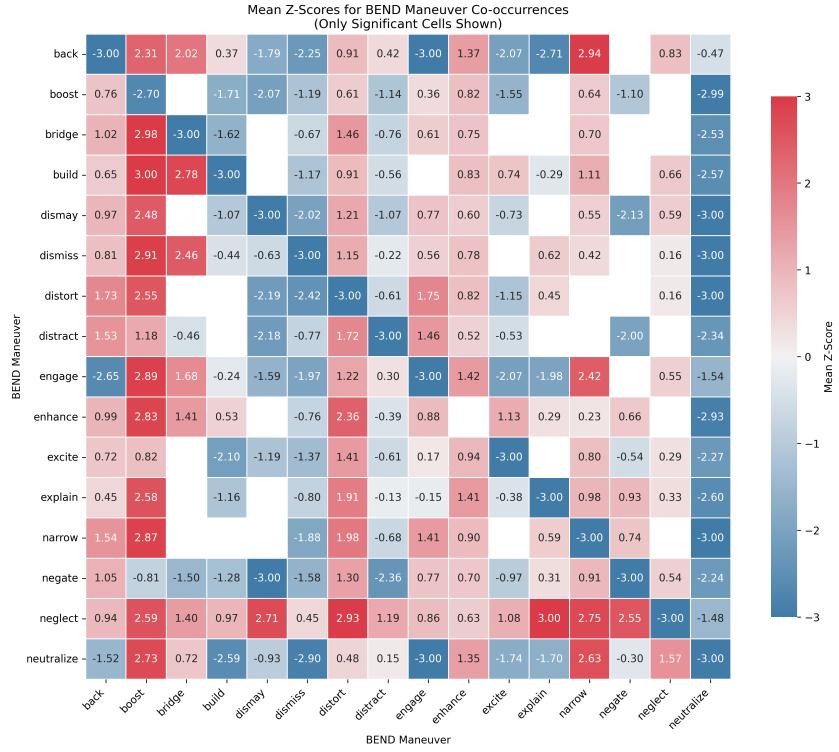


**Fig. 8** Co-occurrence heatmap of the Balikatan 22 data - note that there is no concept of intended vs detected BEND maneuvers in real data

being requested for this particular message - with the BEND maneuver requests specific to the message. However, placing the definitions in the system prompt - increasing the distance between overall BEND definitions and the specific ones being requested - significantly increases the accuracy of the LLM returned responses. However, the methodology we take for trying to re-request the LLM for a better BEND response provided no significant improvement. Finally, the statistically significant prevalence of negative z-scores indicates that the LLM - based on our prompts - is under-representing BEND maneuvers on average.

## 6 Validation

Additional validation comes from directly comparing network metrics between the outputs of SynX + AESOP scenario and the Balikatan 22 dataset. We evaluated SynX outputs on five different network metrics. Three were suggested by Chang et al. in 2024 during their development of an LLM-based synthetic network simulation.[9] We

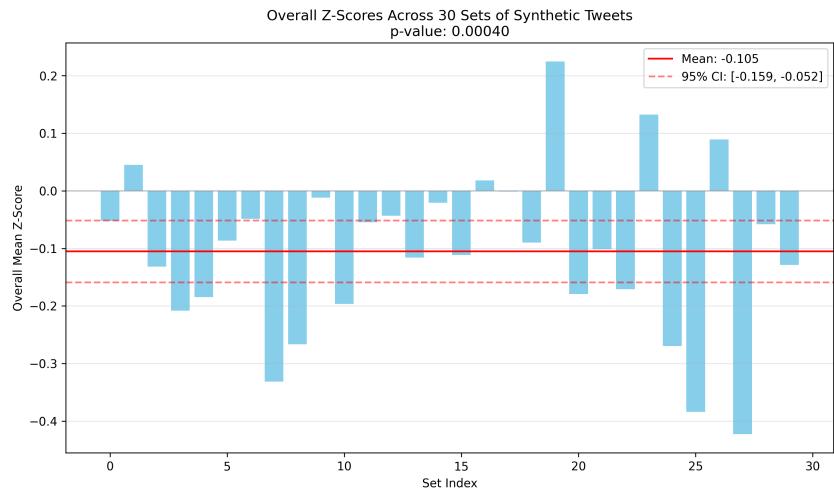


**Fig. 9** Co-occurrence matrix across all runs that had the BEND definitions in the system prompt but did not repoll the LLM

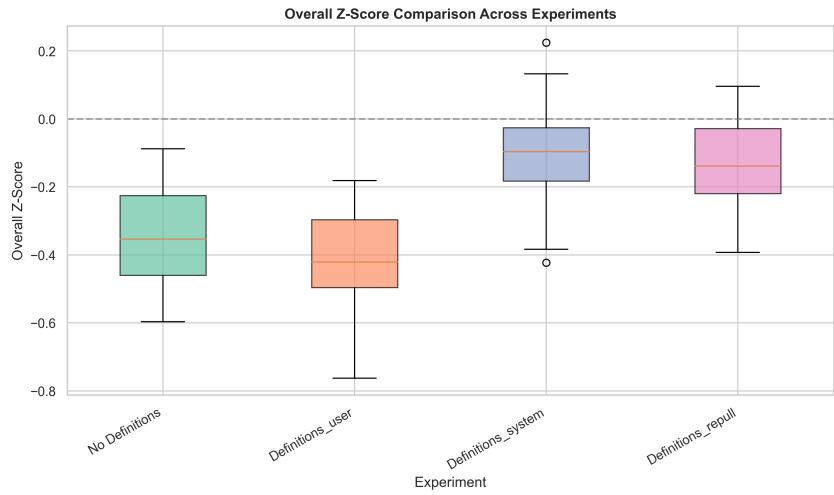
added Degree Distribution of the Tweet x Tweet (Retweet) network because ORA-Pro's primary method for stance propagation depends upon this network being well-connected. It is not necessary that SynX be precisely equivalent to the Balikatan 22 scenario but that the differences should be plausible. We are more concerned with matching shape and form than exact values. The four metrics are:

- Average Shortest Path of Agent x Agent (All Communication)
- Proportion of nodes in the Largest Connected Component of Agent x Agent (All Communication)
- Modularity of Agent x Agent (All Communication)
- Degree Distribution of Tweet x Tweet (Retweet)

From the results in Table 8 we can see that SynX falls within acceptable ranges (within 10%) for two of the three Agent x Agent network metrics when compared with



**Fig. 10** A comparison of the overall distance from the mean for all runs that had the BEND definitions in the system prompt but did not repoll the LLM



**Fig. 11** A comparison of the overall distance from the mean across all BEND maneuvers from each of the four major experiment setups

**Table 7** Stylized Facts

Summary	Effect	Source
Attentions spans limit how many users are affected	Reach is determined by leadership	Kang and Lerman, 2013. [23] Lu et al., 2014. [24]
Commonality of reposts	High level of derivative messages (50+%)	Beskow and Carley's "Agent Based Simulation of Bot Disinformation Maneuvers in Twitter" from 2019.[25]
Which messages are being reposted	Power law distribution	Lu et al., 2014. [24]
Real world likelihood of BEND maneuvers	variable-order Markov model with set-based context and non-replacement sampling based on tweets covering the Balikatan 22 exercise in APR 2022	Lepird, 2024.[22]

**Table 8** Network Metrics Comparison for Agent x Agent Network

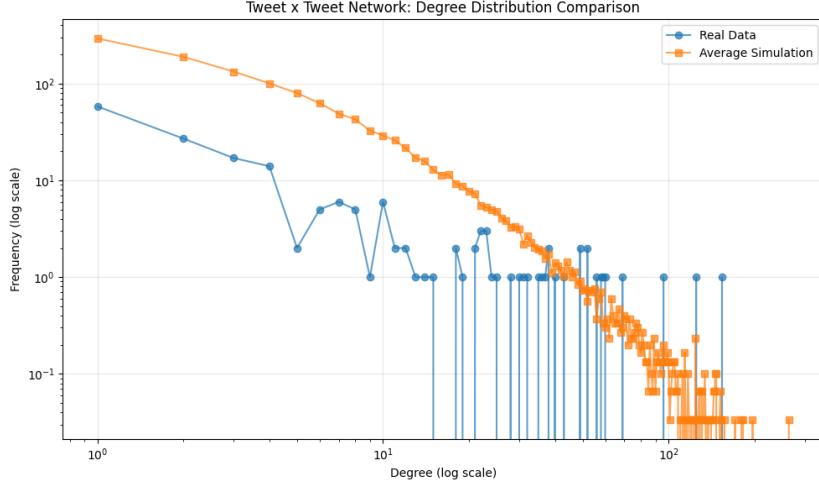
Metric	Real	Sim Mean	Sim Std	Abs. Diff.	Perc. Diff. (%)
Normalized Avg. Shortest Path	0.4158	0.4429	0.0294	-0.0272	-6.54
LCC Proportion	0.8738	0.9583	0.0123	-0.0845	-9.67
Modularity	0.5006	0.9174	0.003	-0.4169	-83.28

**Table 9** Degree Distribution Statistics for Tweet x Tweet Network

Network	Node Count	Mean Degree	Median Degree	Max Degree	Std Dev
Real Data	2230	0.92	0.0	154	6.64
Simulation Avg	10457	0.88	0.0	148	4.74

the Balikatan 22 real dataset. SynX produces data sets that have agents that are too modular when compared to the example set. This is likely due to the constraints placed upon SynX by adhering to the AESOP scenario which has rigidly defined narrative groups. Regardless, more work will need to be done to improve AESOP and SynX in this area.

Finally, when looking at the Tweet x Tweet (Retweet) network results we can see that Synx faithfully recreates a scale-free network and closely mimics the Balikatan network - see Table 9 and Figs. 12 and 13.



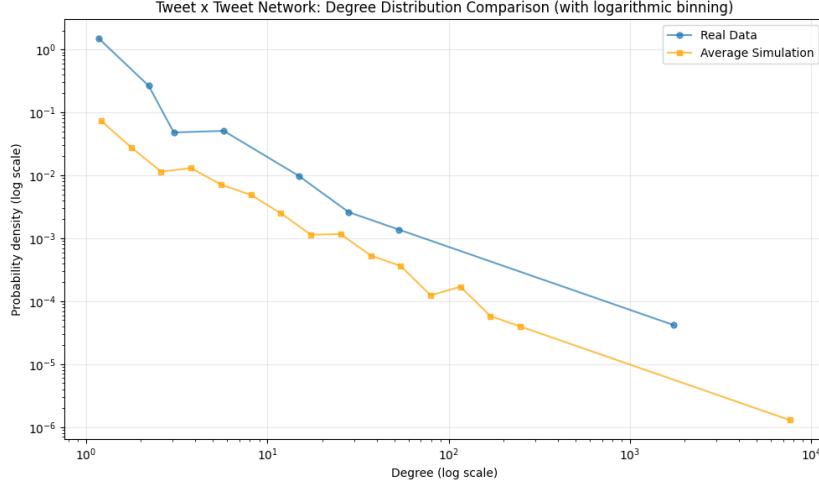
**Fig. 12** Both the real data and the SynX display appropriate power-law properties for the Tweet x Tweet (Retweet) network

## 7 Future Work and Limitations

There are significant limitations to the current model that can be corrected in future work. First, the agents in the current model have opinions matrices that map their stance on each topic in the scenario; however, there is currently no population modeling implemented to feed these matrices nor is there information diffusion implemented such that these stances change over time. Thus, while the actors respond in accordance with their opinions, the simulation is currently a perfect model of arguing on Internet forums - no one ever changes their mind.

Furthermore, more validation should be done. If the goal is simply to train on BEND maneuver detection and response, then the current validation network metrics and BEND evaluations show that SynX is sufficient. However, training and instruction on other aspects of social networks might require additional, unevaluated, network properties. This includes more research into matching the Agent x Agent modularity against real world datasets. Netanomics does have a fully featured evaluation tool for synthetic social media data sets, MOMUS. MOMUS will verify the synthetic dataset's adherence to the AESOP scenario, the semantic and syntactic content of the messages, and the overall network structure against real world data. This will provide a more thorough evaluation of the dataset.

Additional work should be done to improve the re-request methodology from the LLM. We provided the same prompt to the LLM again but added the original returned results as well as an explanation of what was present and what was missing. Other



**Fig. 13** Binning is not always a good idea, but with the discontinuous data from the real data this chart helps depict the form convergence between the real and synthetic data sets

methodologies should be explored, including asking the LLM itself to improve the message.

Finally, work is also required within the BEND evaluations. While the AESOP scenario comes with images and news sites, the current ORA-Pro and NetMapper BEND evaluation is done only on the meta-data and the text. Future detection should include an evaluation of the included images or referenced URLs. Current detection is also focused on deriving the intention of the author of a message rather than measuring the actual effects the message has on the narratives and networks in the corpus. Future studies should include effectiveness measures evaluated against real datasets.

## 8 Conclusion

We create an agent-based simulation, SynX, to model synthetic social media based on input scenario templates from Netanomics’ AESOP tool. Our validation testing shows that SynX is capable of creating highly dynamic datasets that present BEND maneuvers similar to those of real world datasets. Additionally, we experimented with different techniques for interaction with an LLM and found that the inclusion of BEND definitions in system prompts was most effective. The validation and results show that SynX provides an effective way to create tailored social media datasets for analyst training on BEND maneuvers.

**Acknowledgments.** The research for this paper was supported in part by the US Army project Scalable Technologies for Social Cybersecurity (W911NF20D0002) and

by the Office of Naval Research under project Minerva-Multi-Level Models of Covert Online Information Campaigns (N000142112765) and by the center for Informed Democracy and Social-cybersecurity (IDeAS) and the center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research, Army, or the US Government.

## References

- [1] Badawy, A., Addawood, A., Lerman, K., Ferrara, E.: Characterizing the 2016 russian IRA influence campaign **9**(1), 31 <https://doi.org/10.1007/s13278-019-0578-6> . Accessed 2024-04-19
- [2] Carley, K.M.: Social cybersecurity: An emerging science. Computational and Mathematical Organization Theory **26**(4), 365–381 (2020) <https://doi.org/10.1007/s10588-020-09322-9>
- [3] National Academies of Sciences, Engineering, and Medicine: A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis. The National Academies Press, Washington, DC (2019). <https://doi.org/10.17226/25335> . See chapter 6
- [4] Beskow, D.M., Carley, K.M.: Social cybersecurity: an emerging national security requirement **99**(2), 117–127. Accessed 2024-04-18
- [5] Blane, J.T., Bellutta, D., Carley, K.M.: Social-cyber maneuvers during the COVID-19 vaccine initial rollout: Content analysis of tweets **24**(3), 34040 <https://doi.org/10.2196/34040> . Accessed 2024-04-18
- [6] Meadows, D.L., Behrens, W.W., Meadows, D.H., Naill, R.F., Randers, J., Zahn, E.: Dynamics of Growth in a Finite World. Wright-Allen Press, Cambridge, MA (1974). [https://elmoukrie.com/wp-content/uploads/2020/12/dynamics-of-growth-in-a-finite-world\\_nodrm.pdf](https://elmoukrie.com/wp-content/uploads/2020/12/dynamics-of-growth-in-a-finite-world_nodrm.pdf)
- [7] Chopard, B., Droz, M.: Cellular automata: Modelling of physical systems. In: Cellular Automata Modeling of Physical Systems, pp. 6–13. Cambridge University Press, Cambridge, MA (1998). <https://doi.org/10.1017/CBO9780511549755>
- [8] Goldstein, J.: Emergence as a construct: History and issues. Emergence **1**(1), 49–72 (1999) [https://doi.org/10.1207/s15327000em0101\\_4](https://doi.org/10.1207/s15327000em0101_4)
- [9] Chang, S., Chaszczewicz, A., Wang, E., Josifovska, M., Pierson, E., Leskovec, J.: LLMs generate structurally realistic social networks but overestimate political homophily. These authors contributed equally: Serina Chang, Alicja Chaszczewicz (2024). <https://arxiv.org/abs/2408.16629>

- [10] Hämäläinen, P., Tavast, M., Kunnari, A.: Evaluating large language models in generating synthetic hci research data: a case study. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI '23. Association for Computing Machinery, New York, NY, USA (2023). <https://doi.org/10.1145/3544548.3580688> . <https://doi.org/10.1145/3544548.3580688>
- [11] Park, J.S., O'Brien, J.C., Cai, C.J., Morris, M.R., Liang, P., Bernstein, M.S.: Generative agents: Interactive simulacra of human behavior. arXiv preprint arXiv:2304.03442 (2023)
- [12] Veselovsky, V., Ribeiro, M.H., Arora, A., Josifoski, M., Anderson, A., West, R.: Generating faithful synthetic data with large language models: A case study in computational social science. arXiv preprint arXiv:2305.15041 (2023)
- [13] Vongthongsrir, K.: Using LLMs for Synthetic Data Generation: The Definitive Guide. <https://www.confident-ai.com/blog/the-definitive-guide-to-synthetic-data-generation-using-lmms>. Confident AI Blog, Accessed: 2025-03-14 (2023)
- [14] Guo, X., Chen, Y.: Generative ai for synthetic data generation: Methods, challenges and the future. arXiv preprint arXiv:2403.04190 (2024)
- [15] Kulkarni, S., Banerjee, J., Panchumarthy, M.: Generate Synthetic Counterparty (CR) Risk Data with Generative AI Using Amazon Bedrock LLMs and RAG. Amazon Web Services Machine Learning Blog (2025). <https://aws.amazon.com/blogs/machine-learning/generate-synthetic-counterparty-cr-risk-data-with-generative-ai-using-amazon-bedrock-llms-and-rag/>
- [16] Blane, J.T., Moffitt, J.D., Carley, K.M.: Simulating social-cyber maneuvers to deter disinformation campaigns. In: Thomson, R., Hussain, M.N., Dancy, C., Pyke, A. (eds.) Social, Cultural, and Behavioral Modeling vol. 12720, pp. 153–163. Springer. [https://doi.org/10.1007/978-3-030-80387-2\\_15](https://doi.org/10.1007/978-3-030-80387-2_15) . [https://link.springer.com/10.1007/978-3-030-80387-2\\_15](https://link.springer.com/10.1007/978-3-030-80387-2_15) Accessed 2024-04-18
- [17] Dipple, S., Carley, K.M.: Construct User Guide. <https://www.cmu.edu/casos-center/publications/cmu-s3d-23-104.pdf>
- [18] OpenAI: The Official Python Library for the OpenAI API. OpenAI. original-date: 2020-10-25T23:23:54Z. <https://github.com/openai/openai-python> Accessed 2024-04-19
- [19] Carley, K.M.: ORA: A Toolkit for Dynamic Network Analysis and Visualization. In: Alhajj, R., Rokne, J. (eds.) Encyclopedia of Social Network Analysis and Mining. Springer, New York, NY (2017). [https://doi.org/10.1007/978-1-4614-7163-9\\_309-1](https://doi.org/10.1007/978-1-4614-7163-9_309-1)

- [20] Ng, L.H.X., Carley, K.M.: Deflating the chinese balloon: types of twitter bots in us-china balloon incident. EPJ Data Science **12**(1), 63 (2023) <https://doi.org/10.1140/epjds/s13688-023-00440-3>
- [21] Jacobs, C.S., Ng, L.H.X., Carley, K.M.: Tracking china's cross-strait bot networks against taiwan. In: Social, Cultural, and Behavioral Modeling (SBP-BRiMS 2023), pp. 115–125. Springer, New York, NY (2023). [https://doi.org/10.1007/978-3-031-43129-6\\_12](https://doi.org/10.1007/978-3-031-43129-6_12) . [https://link.springer.com/chapter/10.1007/978-3-031-43129-6\\_12](https://link.springer.com/chapter/10.1007/978-3-031-43129-6_12)
- [22] Lepird, C.S.: Pink slime: Measuring, finding, and countering online threats to local news. PhD thesis, Carnegie Mellon University (October 2024). [https://christine-lepird.github.io/Lepird\\_Thesis\\_Oct2024.pdf](https://christine-lepird.github.io/Lepird_Thesis_Oct2024.pdf)
- [23] Kang, J.-H., Lerman, K.: LA-CTR: A limited attention collaborative topic regression for social media **27**(1), 1128–1134 <https://doi.org/10.1609/aaai.v27i1.8451> . Accessed 2024-04-18
- [24] Lu, Y., Zhang, P., Cao, Y., Hu, Y., Guo, L.: On the frequency distribution of retweets **31**, 747–753 <https://doi.org/10.1016/j.procs.2014.05.323> . Accessed 2024-04-18
- [25] Beskow, D.M., Carley, K.M.: Agent based simulation of bot disinformation maneuvers in twitter. In: 2019 Winter Simulation Conference (WSC), pp. 750–761. IEEE. <https://doi.org/10.1109/WSC40007.2019.9004942> . <https://ieeexplore.ieee.org/document/9004942/> Accessed 2024-04-18

# **Appendix C**

## **JSON Schema for Data Standard**

The following is a schema for the AESOP-*c*SynTel/X Data Standard and examples

## JSON SCHEMA

### ACTOR SCHEMA

```
{  
    "$schema": "https://json-schema.org/draft/2020-12/schema",  
    "title": "Actor",  
    "type": "object",  
    "properties": {  
        "aesop_id": {  
            "type": "string",  
            "description": "Unique AESOP ID for the actor."  
        },  
        "name": {  
            "type": "string",  
            "description": "The name of the actor (person, organization, or bot)."  
        },  
        "type": {  
            "type": "string",  
            "description": "Type of actor (e.g., 'Bot', 'Organization', 'Person')."  
        },  
        "leader_type": {  
            "type": "string",  
            "description": "Describes if the actor is a leader in a political, social, or organizational sense."  
        },  
        "title": {  
            "type": "string"  
        },  
        "leads": {  
            "type": "string"  
        },  
        "age": {  
            "type": "string"  
        },  
        "gender": {  
            "type": "string"  
        },  
        "race": {  
            "type": "string"  
        },  
        "nationality": {  
            "type": "string"  
        },  
        "real_person": {  
            "type": "boolean",  
            "description": "Indicates whether this actor represents a real individual or an artificial entity."  
        },  
        "description": {  
            "type": "string"  
        },  
        "bio_prompt": {  
            "type": "string",  
            "description": "Text prompt used to generate an AI biography for the actor."  
        },  
        "ai_bio": {  
            "type": "string",  
            "description": "AI-generated biography for this actor."  
        },  
        "image_prompt": {  
            "type": "string"  
        },  
        "ai_image": {  
            "type": "string"  
        },  
        "day_zero": {  
            "type": "boolean"  
        },  
        "legal": {  
            "type": "string",  
            "description": "Legal status or classification of the actor."  
        }  
    }  
}
```

```
        "description": "Legal disclaimer regarding data ownership and confidentiality."
    },
    "required": [
        "aesop_id",
        "name",
        "type"
    ]
}
```

## ACTOR ORG SCHEMA

```
{  
    "$schema": "https://json-schema.org/draft/2020-12/schema",  
    "title": "Actor",  
    "type": "object",  
    "properties": {  
        "aesop_id": {  
            "type": "string",  
            "description": "Unique AESOP ID for the actor."  
        },  
        "name": {  
            "type": "string",  
            "description": "The name of the actor (person, organization, or bot)."  
        },  
        "type": {  
            "type": "string",  
            "description": "Type of actor (e.g., 'Bot', 'Organization', 'Person')."  
        },  
        "leader_type": {  
            "type": "string",  
            "description": "Describes if the actor is a leader in a political, social, or organizational sense."  
        },  
        "title": {  
            "type": "string"  
        },  
        "leads": {  
            "type": "string"  
        },  
        "age": {  
            "type": "string"  
        },  
        "gender": {  
            "type": "string"  
        },  
        "race": {  
            "type": "string"  
        },  
        "nationality": {  
            "type": "string"  
        },  
        "real_person": {  
            "type": "boolean",  
            "description": "Indicates whether this actor represents a real individual or an artificial entity."  
        },  
        "description": {  
            "type": "string"  
        },  
        "bio_prompt": {  
            "type": "string",  
            "description": "Text prompt used to generate an AI biography for the actor."  
        },  
        "ai_bio": {  
            "type": "string",  
            "description": "AI-generated biography for this actor."  
        },  
        "image_prompt": {  
            "type": "string"  
        },  
        "ai_image": {  
            "type": "string"  
        },  
        "day_zero": {  
            "type": "boolean"  
        },  
        "legal": {  
            "type": "string",  
            "description": "Legal disclaimer regarding data ownership and confidentiality."  
        }  
    }  
}
```

```
},
"required": [
  "aesop_id",
  "name",
  "type"
]
}
```

## ACCOUNT SCHEMA

```
{  
    "$schema": "https://json-schema.org/draft/2020-12/schema",  
    "title": "Account",  
    "type": "object",  
    "properties": {  
        "owner_aesop_id": {  
            "type": "string",  
            "description": "Unique AESOP ID of the actor owning this account."  
        },  
        "twitter_id": {  
            "type": "string",  
            "description": "Unique Twitter ID for this account."  
        },  
        "screen_name": {  
            "type": "string",  
            "description": "Public-facing Twitter username of the account."  
        },  
        "day_zero": {  
            "type": "boolean",  
            "description": "Indicates if this account is known to the Training Audience from the first day of an event."  
        },  
        "user_active_start_date": {  
            "type": "string",  
            "format": "date",  
            "description": "Date when the account activity starts."  
        },  
        "user_active_end_date": {  
            "type": "string",  
            "format": "date",  
            "description": "Date when the account activity ends."  
        },  
        "features": {  
            "type": "array",  
            "items": {  
                "type": "boolean"  
            },  
            "description": "Boolean feature flags indicating various account attributes."  
        },  
        "bend_maneuvers": {  
            "type": "array",  
            "items": {  
                "type": "string"  
            },  
            "description": "BEND maneuvers associated with this account."  
        },  
        "maneuvering_upon": {  
            "type": "string"  
        },  
        "tweet_distribution": {  
            "type": "string",  
            "description": "Pattern of tweets over time (e.g., Day 1: 10%, Day 2: 30%)."  
        },  
        "account_bio": {  
            "type": "string",  
            "description": "Twitter bio of the account."  
        },  
        "account_location": {  
            "type": "string"  
        },  
        "account_url": {  
            "type": "string"  
        },  
        "account_creation_date": {  
            "type": "string",  
            "format": "date"  
        },  
        "number_of_followers": {  
    }
```

```

        "type": "string",
        "description": "Number of followers for this account."
    },
    "number_following": {
        "type": "string"
    },
    "number_of_languages_user_tweets_in": {
        "type": "string"
    },
    "percent_tweets_in_language": {
        "type": "string"
    },
    "num_original_tweets_per_day": {
        "type": "array",
        "items": {
            "type": "integer"
        },
        "description": "Range of original tweets per day."
    },
    "top_topics": {
        "type": "string"
    },
    "num_mentions_per_tweet": {
        "type": "array",
        "items": {
            "type": "integer"
        },
        "description": "Range of mentions per tweet."
    },
    "accounts_to_mention": {
        "type": "string"
    },
    "mention_others": {
        "type": "string"
    },
    "num_retweets_per_day": {
        "type": "array",
        "items": {
            "type": "integer"
        }
    },
    "num_replies_per_day": {
        "type": "array",
        "items": {
            "type": "integer"
        }
    },
    "num_quotes_per_day": {
        "type": "array",
        "items": {
            "type": "integer"
        }
    },
    "retweet_quote_valence": {
        "type": "string"
    },
    "retweet_quote_categories": {
        "type": "string"
    },
    "accounts_to_retweet_quote": {
        "type": "string"
    },
    "other_accounts_retweet_quote": {
        "type": "string"
    },
    "num_hashtags_per_post": {
        "type": "array",
        "items": {
            "type": "integer"
        }
    }
}

```

```

},
"top_hashtags": {
    "type": "string"
},
"other_hashtags": {
    "type": "string"
},
"top_words": {
    "type": "string"
},
"other_words_used": {
    "type": "string"
},
"when_does_account_tweet_gmt": {
    "type": "array",
    "items": {
        "type": "integer"
    }
},
"percent_tweets_pos_neg_neut": {
    "type": "string"
},
"text_ratio": {
    "type": "string"
},
"image_ratio": {
    "type": "string"
},
"video_ratio": {
    "type": "string"
},
"identity_markers": {
    "type": "string"
},
"stylistic_markers": {
    "type": "string"
},
"legal": {
    "type": "string",
    "description": "Legal disclaimer regarding data ownership and confidentiality."
}
},
"required": [
    "owner_aesop_id",
    "twitter_id",
    "screen_name"
]
}

```

## ACCOUNT BOT SCHEMA

```
{  
  "$schema": "https://json-schema.org/draft/2020-12/schema",  
  "title": "Account",  
  "type": "object",  
  "properties": {  
    "owner_aesop_id": {  
      "type": "string",  
      "description": "Unique AESOP ID of the actor owning this account."  
    },  
    "twitter_id": {  
      "type": "string",  
      "description": "Unique Twitter ID for this account."  
    },  
    "screen_name": {  
      "type": "string",  
      "description": "Public-facing Twitter username of the account."  
    },  
    "day_zero": {  
      "type": "boolean",  
      "description": "Indicates if this account is known to the Training Audience from the first day of an event."  
    },  
    "user_active_start_date": {  
      "type": "string",  
      "format": "date",  
      "description": "Date when the account activity starts."  
    },  
    "user_active_end_date": {  
      "type": "string",  
      "format": "date",  
      "description": "Date when the account activity ends."  
    },  
    "features": {  
      "type": "array",  
      "items": {  
        "type": "boolean"  
      },  
      "description": "Boolean feature flags indicating various account attributes."  
    },  
    "bend_maneuvers": {  
      "type": "array",  
      "items": {  
        "type": "string"  
      },  
      "description": "BEND maneuvers associated with this account."  
    },  
    "maneuvering_upon": {  
      "type": "string"  
    },  
    "tweet_distribution": {  
      "type": "string",  
      "description": "Pattern of tweets over time (e.g., Day 1: 10%, Day 2: 30%)."  
    },  
    "account_bio": {  
      "type": "string",  
      "description": "Twitter bio of the account."  
    },  
    "account_location": {  
      "type": "string"  
    },  
    "account_url": {  
      "type": "string"  
    },  
    "account_creation_date": {  
      "type": "string",  
      "format": "date"  
    },  
    "number_of_followers": {  
      "type": "string",  
      "description": "Number of followers."  
    }  
  }  
}
```

```

    "description": "Number of followers for this account."
},
"number_following": {
    "type": "string"
},
"number_of_languages_user_tweets_in": {
    "type": "string"
},
"percent_tweets_in_language": {
    "type": "string"
},
"num_original_tweets_per_day": {
    "type": "array",
    "items": {
        "type": "integer"
    },
    "description": "Range of original tweets per day."
},
"top_topics": {
    "type": "string"
},
"num_mentions_per_tweet": {
    "type": "array",
    "items": {
        "type": "integer"
    },
    "description": "Range of mentions per tweet."
},
"accounts_to_mention": {
    "type": "string"
},
"mention_others": {
    "type": "string"
},
"num_retweets_per_day": {
    "type": "array",
    "items": {
        "type": "integer"
    }
},
"num_replies_per_day": {
    "type": "array",
    "items": {
        "type": "integer"
    }
},
"num_quotes_per_day": {
    "type": "array",
    "items": {
        "type": "integer"
    }
},
"retweet_quote_valence": {
    "type": "string"
},
"retweet_quote_categories": {
    "type": "string"
},
"accounts_to_retweet_quote": {
    "type": "string"
},
"other_accounts_retweet_quote": {
    "type": "string"
},
"num_hashtags_per_post": {
    "type": "array",
    "items": {
        "type": "integer"
    }
}

```

```

"top_hashtags": {
    "type": "string"
},
"other_hashtags": {
    "type": "string"
},
"top_words": {
    "type": "string"
},
"other_words_used": {
    "type": "string"
},
"when_does_account_tweet_gmt": {
    "type": "array",
    "items": {
        "type": "integer"
    }
},
"percent_tweets_pos_neg_neut": {
    "type": "string"
},
"text_ratio": {
    "type": "string"
},
"image_ratio": {
    "type": "string"
},
"video_ratio": {
    "type": "string"
},
"identity_markers": {
    "type": "string"
},
"stylistic_markers": {
    "type": "string"
},
"legal": {
    "type": "string",
    "description": "Legal disclaimer regarding data ownership and confidentiality."
}
},
"required": [
    "owner_aesop_id",
    "twitter_id",
    "screen_name"
]
}

```

## EVENT SCHEMA

```
{  
    "$schema": "https://json-schema.org/draft/2020-12/schema",  
    "title": "Event",  
    "type": "object",  
    "properties": {  
        "aesop_id": {  
            "type": "string",  
            "description": "Unique AESOP ID for the event."  
        },  
        "name": {  
            "type": "string",  
            "description": "Name of the event."  
        },  
        "type": {  
            "type": "string",  
            "description": "Categorization of the event (e.g., 'Political', 'Accident')."  
        },  
        "excite_number": {  
            "type": "integer",  
            "description": "Numeric indicator of how much excitement the event generates."  
        },  
        "active_date_range": {  
            "type": "array",  
            "items": {  
                "type": "string",  
                "format": "date-time"  
            },  
            "description": "Date range in which the event is active."  
        },  
        "summary": {  
            "type": "string",  
            "description": "Short textual summary of the event."  
        },  
        "ai_description": {  
            "type": "string",  
            "description": "AI-generated detailed description of the event."  
        },  
        "legal": {  
            "type": "string",  
            "description": "Legal disclaimer regarding data ownership and confidentiality."  
        }  
    },  
    "required": [  
        "aesop_id",  
        "name",  
        "type",  
        "active_date_range"  
    ]  
}
```

## GROUP SCHEMA

```
{  
    "$schema": "https://json-schema.org/draft/2020-12/schema",  
    "title": "Group",  
    "type": "object",  
    "properties": {  
        "aesop_id": {  
            "type": "string",  
            "description": "Unique AESOP ID for the group."  
        },  
        "name": {  
            "type": "string",  
            "description": "Name of the group."  
        },  
        "leaders": {  
            "type": "array",  
            "items": {  
                "type": "string"  
            },  
            "description": "List of AESOP IDs representing group leaders."  
        },  
        "people": {  
            "type": "array",  
            "items": {  
                "type": "string"  
            },  
            "description": "List of AESOP IDs representing full members of the group."  
        },  
        "people_src": {  
            "type": "array",  
            "items": {  
                "type": "string"  
            },  
            "description": "List of AESOP IDs representing members that this group will source from."  
        },  
        "orgs": {  
            "type": "array",  
            "items": {  
                "type": "string"  
            }  
        },  
        "orgs_src": {  
            "type": "array",  
            "items": {  
                "type": "string"  
            }  
        },  
        "bots": {  
            "type": "array",  
            "items": {  
                "type": "string"  
            },  
            "description": "List of bot AESOP IDs affiliated with this group."  
        },  
        "bots_src": {  
            "type": "array",  
            "items": {  
                "type": "string"  
            }  
        },  
        "chatters": {  
            "type": "array",  
            "items": {  
                "type": "string"  
            }  
        },  
        "chatters_src": {  
            "type": "array",  
            "items": {  
                "type": "string"  
            }  
        }  
    }  
}
```

```
        "type": "string"
    }
},
"groups": {
    "type": "array",
    "items": {
        "type": "string"
    }
},
"groups_src": {
    "type": "array",
    "items": {
        "type": "string"
    },
    "description": "List of AESOP IDs representing groups that this group will source from."
},
"legal": {
    "type": "string",
    "description": "Legal disclaimer regarding data ownership and confidentiality."
}
},
"required": [
    "aesop_id",
    "name"
]
}
```

## TOPIC SCHEMA

```
{  
    "$schema": "https://json-schema.org/draft/2020-12/schema",  
    "title": "Topic",  
    "type": "object",  
    "properties": {  
        "aesop_id": {  
            "type": "string",  
            "description": "Unique AESOP ID for the topic."  
        },  
        "topic_name": {  
            "type": "string",  
            "description": "Name of the topic."  
        },  
        "topic_description": {  
            "type": "string",  
            "description": "Brief description of the topic."  
        },  
        "number_of_narratives": {  
            "type": "integer",  
            "description": "Number of narratives associated with this topic."  
        },  
        "narratives": {  
            "type": "array",  
            "items": {  
                "type": "object",  
                "properties": {  
                    "aesop_id": {  
                        "type": "string",  
                        "description": "Unique AESOP ID for the narrative."  
                    },  
                    "narrative_name": {  
                        "type": "string",  
                        "description": "Name of the narrative."  
                    },  
                    "narrative_date_start": {  
                        "type": "string",  
                        "format": "date-time",  
                        "description": "Start date of the narrative."  
                    },  
                    "narrative_date_end": {  
                        "type": "string",  
                        "format": "date-time",  
                        "description": "End date of the narrative."  
                    },  
                    "narrative_stance": {  
                        "type": "string",  
                        "description": "Stance on the topic (e.g., 'Pro', 'Anti', 'Neutral')."  
                    },  
                    "narrative_description": {  
                        "type": "string",  
                        "description": "Description of the narrative."  
                    },  
                    "narrative_group": {  
                        "type": "array",  
                        "items": {  
                            "type": "string"  
                        },  
                        "description": "AESOP IDs of groups associated with this narrative."  
                    },  
                    "narrative_event": {  
                        "type": "array",  
                        "items": {  
                            "type": "string"  
                        },  
                        "description": "AESOP IDs of events linked to this narrative."  
                    },  
                    "narrative_hashtags": {  
                        "type": "array",  
                        "description": "Hashtags associated with this narrative."  
                    }  
                }  
            }  
        }  
    }  
}
```

```
        "items": {
            "type": "string"
        },
        "description": "Hashtags associated with this narrative."
    },
    "bend_maneuvers": {
        "type": "array",
        "items": {
            "type": "string"
        },
        "description": "BEND maneuvers related to this narrative."
    }
},
"required": [
    "aesop_id",
    "narrative_name",
    "narrative_date_start",
    "narrative_date_end"
],
"description": "List of narratives under this topic."
},
"legal": {
    "type": "string",
    "description": "Legal disclaimer regarding data ownership and confidentiality."
}
},
"required": [
    "aesop_id",
    "topic_name",
    "number_of_narratives"
]
}
```

## JSON EXAMPLES

### ACTOR PERSON EXAMPLE

```
{  
    "aesop_id": "3916147220",  
    "name": "Bob J",  
    "type": "Organizational",  
    "title": "President",  
    "leads": "Dog Enthusiasts of America",  
    "age": "25",  
    "gender": "Male",  
    "race": "Caucasian",  
    "nationality": "American",  
    "real_person": false,  
    "entourage_size": 0,  
    "entourage": [],  
    "bio_prompt": "Please provide a character background for Bob J a 25 year-old Male, who is the President of Dog Enthusiasts of America in a realistic roleplaying game. Bob loves dogs and wants everyone else to love them as well.",  
    "ai_bio": "Character Background: Bob J\nName: Bob J\nAge: 25\nGender: Male\nRole: President of Dog Enthusiasts of America\nBackground:\nBob J is a passionate and enthusiastic 25-year-old man with a deep love for dogs. Growing up in a small rural town, Bob always had a special connection with animals, especially dogs. He was raised in a household that instilled in him a sense of compassion and responsibility towards animals from an early age. Bob's parents encouraged his love for dogs by adopting several rescue animals, including dogs, which further solidified his bond with them.\nAfter graduating from high school, Bob pursued a degree in Animal Sciences, with a particular focus on canine behavior and training. He spent his college years actively working with local animal shelters, volunteering his time to help rehabilitate and find homes for abandoned dogs. Bob's dedication and hands-on experience allowed him to develop a deep understanding and empathy for these animals.\nDuring his college years, Bob also started the Dog Enthusiasts Club, a community-driven group focused on advocating for responsible pet ownership, supporting local animal shelters, and promoting dog welfare. This club aimed to educate the public about the benefits of owning dogs while also addressing the importance of proper training, nutrition, and healthcare.\nFollowing his graduation, Bob's passion for dogs led him to take on the role of President of Dog Enthusiasts of America, a national organization aiming to bridge the gap between dog owners, enthusiasts, and professionals across the country. As President, Bob spearheads campaigns and initiatives that promote awareness, education, and responsible ownership.\nBob is known for his unwavering dedication and infectious enthusiasm for his cause. He regularly organizes adoption drives and fundraising events, leveraging both his extensive network and social media presence. Bob believes that fostering a love for dogs amongst others will create a positive impact not only on the lives of these animals but also on society as a whole.\nOn a personal level, Bob's own love for dogs is reflected in his own household. He has a loyal and affectionate dog named Max, an adopted mixed breed. Bob and Max share a deep bond and are often seen accompanying each other on adventures, from exploring nature trails to participating in dog-friendly events.\nThrough his role as President, Bob hopes to inspire and mobilize others to recognize the importance of dogs in our daily lives. He believes that, with the right support, dogs can serve as sources of companionship, emotional well-being, and personal growth. Bob's ultimate goal is to create a world where every dog is loved, cherished, and cared for in a responsible and compassionate manner.",  
    "image_prompt": "",  
    "ai_image": "no_image.png",  
    "description": "Bob loves dogs and wants everyone else to love them as well.",  
    "day_zero": false,  
    "ATM": {},  
    "legal": "All material contained is confidential and proprietary to Netanomics and the CASOS center at CMU."  
}
```

## ACTOR ORG EXAMPLE

```
{  
    "aesop_id": "3362094714",  
    "name": "CoolCats",  
    "type": "News",  
    "organizational_leader": "",  
    "main_account_handle": "",  
    "other_associated_handles": "",  
    "region_targeted": [],  
    "news_bias": "Extreme Left",  
    "organizational_credibility": "High",  
    "questionable_characteristics": [],  
    "bio_prompt": "Please provide several paragraphs of the organizational biography and history for CoolCats.\nIt is an organization that:\nCoolCats is a media company that is only interested in driving the cat narrative forward.\nAdditional descriptors include:\n",  
    "ai_bio": "innovative, forward-thinking, creative, and passionate.\nCoolCats was founded in 2010 by a group of cat enthusiasts with a shared vision of promoting all things related to cats and felines. The organization was born out of a desire to not only celebrate the beloved companionship that cats provide but also to create a platform that showcases their unique qualities and stories. From the beginning, CoolCats set out to be a media company that focuses solely on driving the cat narrative forward and captivating audiences worldwide.\n\nWith a commitment to innovation, CoolCats strives to constantly push boundaries and challenge conventional norms when it comes to cat-related content. The organization prides itself on being at the forefront of digital media and utilizing cutting-edge technologies to deliver engaging and entertaining cat-centric experiences. Through the use of immersive storytelling techniques, original videos, captivating photos, and interactive digital platforms, CoolCats seeks to captivate audiences and provide a fresh perspective on the feline world.\n\nSince its inception, CoolCats has been recognized as a pioneering force in the cat media landscape. The organization has successfully collaborated with renowned photographers, filmmakers, and artists to produce visually stunning content that showcases the beauty, agility, and playfulness of cats. CoolCats has also established strong relationships with cat-related brands, shelters, and organizations, serving as a bridge between the various stakeholders in the cat community.\n\nCoolCats' commitment to creativity and passion is evident in its diverse range of projects. The organization has launched successful online campaigns in support of animal welfare causes, education initiatives, and adoption programs. Additionally, CoolCats has ventured into publishing, producing high-quality magazines and books that explore the allure and mysteries of the feline world. These publications showcase captivating stories, share valuable knowledge, and celebrate the unique bond between humans and cats.\n\nAs a forward-thinking organization, CoolCats understands the power of social media and actively leverages platforms such as Instagram, YouTube, and Facebook to reach and engage with a global audience. Through captivating posts, interactive features, and community-building efforts, CoolCats continues to cultivate a strong online presence, fostering connections among cat lovers and enthusiasts worldwide.\n\nIn summary, CoolCats is an innovative and creative media company that has been dedicated to driving the cat narrative forward since its inception. With a passion for cats and a commitment to pushing boundaries, CoolCats produces captivating content, hosts engaging experiences, and supports various initiatives that celebrate the feline world. Through its unique storytelling techniques and cutting-edge digital platforms, CoolCats seeks to captivate audiences, connect cat lovers, and advocate for the welfare and well-being of cats everywhere.",  
    "image_prompt": "",  
    "ai_image": "no_image.png",  
    "description": "CoolCats is a media company that is only interested in driving the cat narrative forward.",  
    "day_zero": false,  
    "ATM": {},  
    "legal": "All material contained is confidential and proprietary to Netanomics and the CASOS center at CMU."  
}
```

## X/TWITTER ACCOUNT EXAMPLE

```
{  
    "owner_aesop_id": "3362094714",  
    "twitter_id": "3036913001664145858",  
    "screen_name": "CoolCats",  
    "day_zero": false,  
    "user_active_start_date": "2024-05-24",  
    "user_active_end_date": "2024-05-26",  
    "features": [  
        false,  
        false  
    ],  
    "bend_maneuvers": [  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0",  
        "0"  
    ],  
    "maneuvering_upon": "",  
    "tweet_distribution": "Day 1(10%), Day 2(10%), Day 3(10%), Day 4(40%), Day 5(30%)",  
    "account_bio": "",  
    "account_location": "",  
    "account_url": "",  
    "account_creation_date": "2024-12-05",  
    "number_of_followers": "10000",  
    "number_following": "100",  
    "number_of_languages_user_tweets_in": "1",  
    "percent_tweets_in_language": "en:100.00%",  
    "num_original_tweets_per_day": [  
        2,  
        5  
    ],  
    "top_topics": "",  
    "num_mentions_per_tweet": [  
        0,  
        4  
    ],  
    "accounts_to_mention": "",  
    "mention_others": "Yes",  
    "num_retweets_per_day": [  
        4,  
        10  
    ],  
    "num_replies_per_day": [  
        4,  
        10  
    ],  
    "num_quotes_per_day": [  
        4,  
        10  
    ],  
    "retweet_quote_valence": "Mostly Positive",  
    "retweet_quote_categories": "News (60%); Gov (30%); Health (7%); Other(3%)",  
    "accounts_to_retweet_quote": "",  
    "other_accounts_retweet_quote": "Yes",  
    "num_hashtags_per_post": [  
        0,
```

```
        4
    ],
    "top_hashtags": "",
    "other_hashtags": "Yes",
    "top_words": "",
    "other_words_used": "Yes",
    "when_does_account_tweet_gmt": [
        7,
        20
    ],
    "percent_tweets_pos_neg_neut": "Positive (50%), Negative (25%), Neutral (25%)",
    "text_ratio": "1",
    "image_ratio": "0",
    "video_ratio": "0",
    "identity_markers": "None",
    "stylistic_markers": "None",
    "legal": "All material contained is confidential and proprietary to Netanomics and the CASOS center at CMU."
}
```

## EVENT EXAMPLE

```
{  
    "aesop_id": "5840946919",  
    "name": "Accidental death of a cat",  
    "type": "Collision/Accident",  
    "which_objects_involved": "",  
    "who_owns_objects": "",  
    "name_of_objects": "",  
    "excite_number": 3,  
    "active_date_range": [  
        "2024-05-25 11:00",  
        "2024-05-25 13:00"  
    ],  
    "other_countries_involved": [],  
    "cities": "",  
    "other_areas": "",  
    "event_leader": "",  
    "leader_job_title": "",  
    "leader_home_country": [],  
    "positive_hashtags": "",  
    "negative_hashtags": "",  
    "summary": "A cat was accidentally run over by a car in Ramses Park.",  
    "prompt": "Here is some information about Accidental death of a cat, a(n) Collision/Accident:\nThis accident involves a cat and a car. It takes place from 2024-05-25 to 2024-05-25. The purpose/summary of the event is as follows: \nA cat was accidentally run over by a car in Ramses Park.\nGenerate several paragraphs describing the event in detail.",  
    "ai_description": "On May 25, 2024, a tragic accident occurred in Ramses Park involving a cat and a car. It was an ordinary day, with people enjoying the park's serene atmosphere, unaware of the heartbreak events that were about to unfold. The unsuspecting cat, going about its day, crossed paths with an unfortunate vehicle.\nAs fate would have it, the car, driven by an unaware and unsuspecting driver, approached the unsuspecting feline. In a split second, the cat found itself in the wrong place at the wrong time, unable to escape the impending danger. The collision was sudden and unavoidable, causing a wave of panic and distress amongst onlookers who witnessed the heartbreaking scene.\nThe impact of the car against the innocent creature resulted in immediate tragedy. The force was enough to cause serious injuries that led to the untimely demise of the cat. An overwhelming sense of sorrow overcame those who witnessed the unfortunate event, as they realized there was nothing that could have been done to prevent it.\nMembers of the community, park-goers, and animal lovers came together, heartbroken by the unfortunate accident. Some attempted to console and provide support to the distraught driver, understanding that this was an unintentional incident. Others grieved for the loss of an innocent life, reflecting on the fragility and unpredictability of existence.\nOfficials were promptly contacted, and an investigation was initiated to determine the exact details and circumstances surrounding the collision. Statements from witnesses were obtained, reconstructing the event to understand the sequence of events leading to the tragic accident. The incident served as a stark reminder of the importance of attentiveness and cautious driving, even in seemingly safe environments.\nIn the aftermath of the accident, the community rallied together to create awareness about road safety, emphasizing the need for drivers to be ever watchful for animals that may unexpectedly cross their path. This unfortunate event served as a crucial reminder to drivers to exercise heightened caution, particularly in areas where animals are known to roam freely.\nAlthough the accidental death of a cat may appear inconsequential to some, its impact resonates deeply within the hearts of those who value all forms of life. The incident highlights the importance of fostering a community that not only cares for its human members but also extends this compassion to the welfare and safety of animals. It serves as a reminder to us all to be mindful of our surroundings, acknowledging that our actions have consequences that extend beyond our own lives.",  
    "number_of_injects": 0,  
    "injects": [],  
    "ai_image": "no_image.png",  
    "legal": "All material contained is confidential and proprietary to Netanomics and the CASOS center at CMU."  
}
```

## GROUP EXAMPLE

```
{  
  "aesop_id": "3969792916",  
  "name": "Dogs_Good",  
  "group_ratio": "1",  
  "leaders": [  
    "4002506303"  
,  
  "people": [  
    "3916147220",  
    "7879212436"  
,  
  "people_src": [],  
  "orgs": [  
    "4002506303"  
,  
  "orgs_src": [],  
  "bots": [  
    "8590922004"  
,  
  "bots_src": [],  
  "chatters": [],  
  "chatters_src": [],  
  "groups": [],  
  "groups_src": [  
    "7704052942"  
,  
  "GTM": {},  
  "GTT": {},  
  "legal": "All material contained is confidential and proprietary to Netanomics and the CASOS center at CMU."  
}
```

## TOPIC EXAMPLE

```
{
  "aesop_id": "6262843952",
  "topic_name": "Cats",
  "topic_hashtags": [],
  "topic_description": "Cats rule",
  "number_of_narratives": 4,
  "narratives": [
    {
      "aesop_id": "8350793621",
      "narrative_name": "Cat allergens",
      "narrative_date_start": "2024-05-07 00:00",
      "narrative_date_end": "2024-05-28 00:00",
      "narrative_stance": "Anti",
      "narrative_ratio": "1",
      "narrative_description": "Cats do not make the best pets because many people are allergic to them.",
      "narrative_example": "",
      "narrative_group": [
        "3969792916"
      ],
      "narrative_event": [],
      "narrative_hashtags": [
        "nocats",
        "allergies"
      ],
      "bend_maneuvers": [
        "100",
        "20",
        "20",
        "10",
        "10",
        "10",
        "10",
        "10",
        "5",
        "5",
        "5",
        "5",
        "5",
        "5",
        "5",
        "5"
      ]
    },
    {
      "aesop_id": "7946016526",
      "narrative_name": "Cats are the best pets",
      "narrative_date_start": "2024-05-14 00:00",
      "narrative_date_end": "2024-05-26 00:00",
      "narrative_stance": "Pro",
      "narrative_ratio": "1",
      "narrative_description": "Cats make the best pets",
      "narrative_example": "",
      "narrative_group": [
        "8984124093"
      ],
      "narrative_event": [],
      "narrative_hashtags": [
        "cats",
        "pets"
      ],
      "bend_maneuvers": [
        "100",
        "20",
        "20",
        "10",
        "10",
        "10"
      ]
    }
  ]
}
```

```

        "10",
        "10",
        "5",
        "5",
        "5",
        "5",
        "5",
        "5",
        "5"
    ],
},
{
    "aesop_id": "4076237060",
    "narrative_name": "Death of a cat",
    "narrative_date_start": "2024-05-25 00:00",
    "narrative_date_end": "2024-05-25 03:00",
    "narrative_stance": "Neutral",
    "narrative_ratio": "1",
    "narrative_description": "A cat was accidentally killed today.",
    "narrative_example": "",
    "narrative_group": [
        "7704052942",
        "8984124093"
    ],
    "narrative_event": [
        "5840946919"
    ],
    "narrative_hashtags": [
        "petnews"
    ],
    "bend_maneuvers": [
        "100",
        "20",
        "20",
        "10",
        "10",
        "10",
        "10",
        "10",
        "5",
        "5",
        "5",
        "5",
        "5",
        "5",
        "5",
        "5"
    ],
},
{
    "aesop_id": "9566229720",
    "narrative_name": "Cat Death",
    "narrative_date_start": "2024-05-25 00:00",
    "narrative_date_end": "2024-05-25 04:00",
    "narrative_stance": "Pro",
    "narrative_ratio": "1",
    "narrative_description": "A great tragedy happened today as a cat died.",
    "narrative_example": "",
    "narrative_group": [
        "8984124093"
    ],
    "narrative_event": [
        "5840946919"
    ],
    "narrative_hashtags": [
        "cats",
        "tragedy"
    ],
    "bend_maneuvers": [

```

```
"100",
"20",
"20",
"10",
"10",
"10",
"10",
"10",
"10",
"5",
"5",
"5",
"5",
"5",
"5",
"5",
"5"
    ]
},
],
"legal": "All material contained is confidential and proprietary to Netanomics and the CASOS center at CMU."
}
```

## **Appendix D**

# **Large Language Models and Synthetic Social Media Generation**

An overview of LLMs and Synthetic Generation with an emphasis on prompts and data evaluation.

# Large Language Models and Synthetic Social Media Generation

Matthew Hicks



Carnegie Mellon University



1

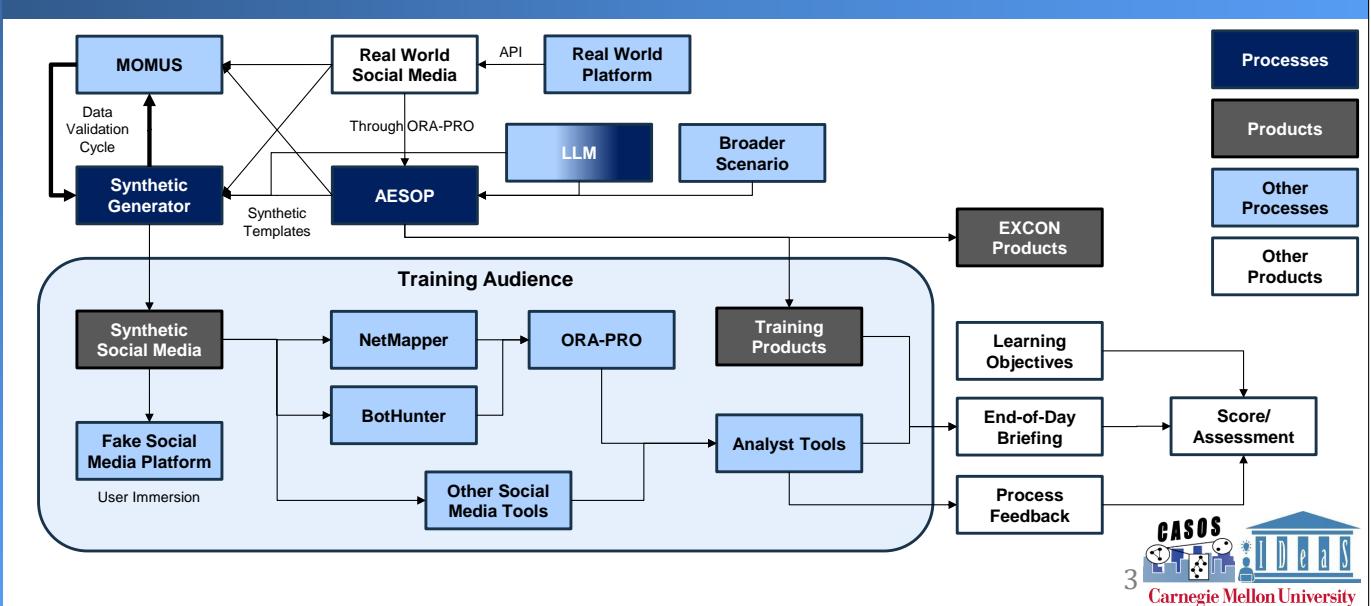
## Outline

- ❑ An Introduction to Realistic Training
- ❑ AI-Enabled Scenario Orchestration and Planning (AESOP) tool
- ❑ SynX and SynTel- Synthetic Generation for X/Twitter and Telegram
- ❑ LLMs
- ❑ Data Validation
- ❑ MOMUS

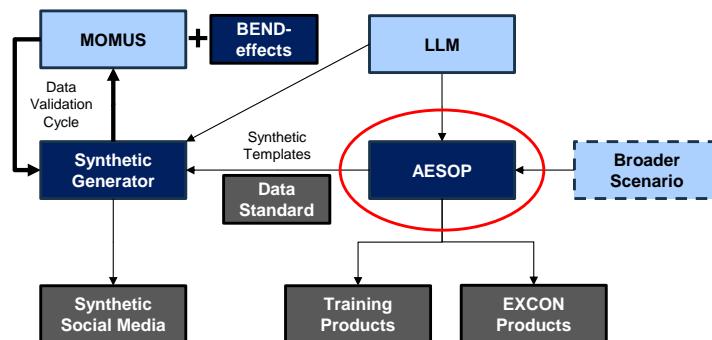


2

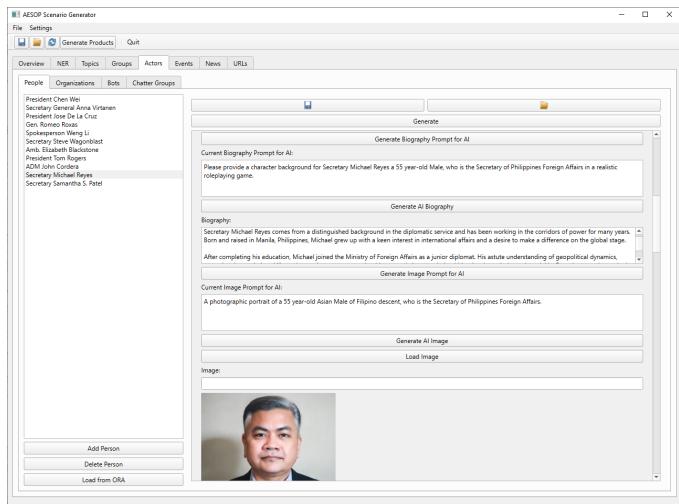
## Why? Realistic Training.



## Synthetic Data Generation



# AI-Enabled Scenario Orchestration and Planning (AESOP)



The Scenario Generator provides scaffolding for developing a range of templates to enable synthetic data generation: Actors, Topics, Groups, Events, News, URLs, etc.

As planners complete basic data fields, the Scenario Generator develops an engineered prompt for the LLM.

Planners can make additional changes to the prompt as required.

Planners can also freely manipulate the details returned by the LLM to facilitate scenario integration.



# AI-Enabled Scenario Orchestration and Planning (AESOP)

## AESOP successfully leverages LLMs for help with:

Biographies for people and organizations

Press Releases

Military Operations Orders

Intelligence Summaries

News Articles

Photographs for accounts – sometimes with help

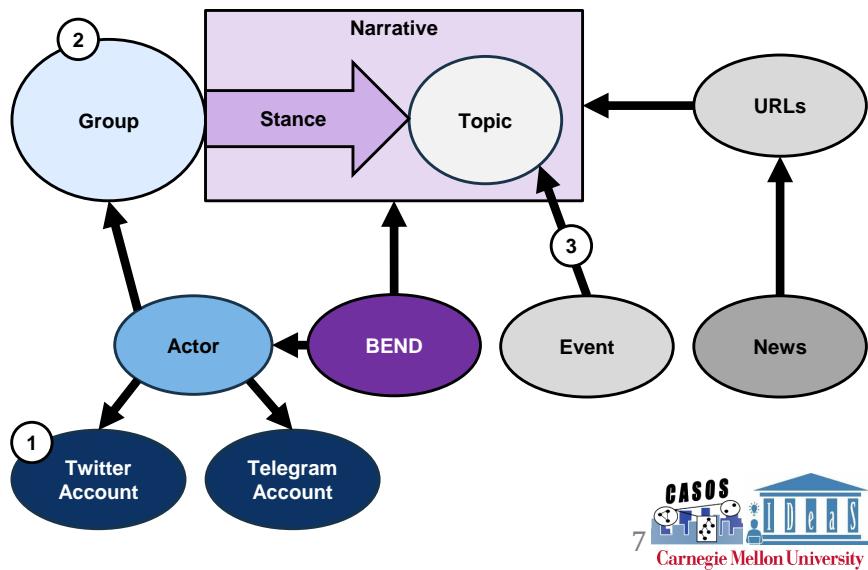
Photographs for news – with help

Memes – with help



# AESOP Data Structure

- 1 We are specifying Twitter Account information
- 2 We can also specify a Group by ensuring the owning actor for an account is a member of it. Group membership is split in two:
  - a) Full members – full members are eligible to be retweeted by other full members and are also eligible to originally tweet about Group associated Narratives
  - b) Source members – source members are eligible to be retweeted by full members but will never originally tweet about Group associated Narratives
- 3 Groups have associated Narratives that they espouse (which include BEND maneuver likelihood). These Narratives, in-turn, are associated with Events. Events will cause spikes in Narrative tweets, etc.



## AESOP Outputs

### Scenario Description

#### Balikatan 2045

##### General Info

Date Range: 2045-04-01 to 2045-05-01

##### Countries of Interest:

Philippines, United States

##### Regions of Interest:

South-eastern Asia

##### Other Locations of Interest:

##### Scenario Description

Balikatan 2045 was an exercise run by the Philippines and United States that took place from April 1 to May 1 in 2045. Aimed at developing and preserving the security relationship between the two countries, participants were trained in crisis action planning, counter-terrorism operations and increasing the interoperability and coordination of their forces.

However, the exercise was overshadowed by a severe health crisis in the region, most notably the outbreak of Typhoon Hayman. This natural disaster, which first struck the Philippines in 2045 and its rapid spread to countries such as the Philippines, Australia, China, Japan, Thailand, and Malaysia, had a huge impact on the medical community. Symptoms of the virus include fever, chills, nausea, diarrhea, muscle aches and pains, and with no vaccine currently available, preventive measures are the best way to contain the spread.

Tragedy was also forced during Balikatan 2045, as the USS Garlie collided with the Liberian-flagged tanker Casinos MC in the Strait of Malacca. The US Navy warship was conducting a routine patrol when it came into contact with the cargo ship. The collision led to various compartments of the destroyer. Ten US Navy sailors were killed in the incident, and rescue efforts were launched by various agencies, including the Singapore Coast Guard, Royal Malaysian Navy, and the US Navy. Unfortunately, no further survivors were located and the cause of the collision is still under investigation.

Finally, Balikatan 2045 was marked by Typhoon Hayman. This Climate Event resulted in extensive damage to vital infrastructure in the Philippines, leaving thousands without food, water, or medical care. In response, the Red Cross activated their Cluster System, consisting of a collection of different humanitarian organizations. The U.S. Navy, along with other governments and NGOs, began providing relief to the hardest hit areas in the form of food, medical supplies and personnel. Volunteers were organized to distribute blankets and

### Actor Information Cards

#### Chandu Claver

##### General Info



Name: Chandu Claver  
Leader Type: Political  
Title: Leader  
Organization: Cordillera's People Alliance  
Gender: Male  
Age: 56  
Race: Asian  
Nationality: Filipino

##### Biography

Chandu Claver is a 56 year-old man from Cordillera, a rural region that has been generally overlooked by the rest of the country. Having grown up in this hamlet among its close-knit community, Chandu has grown to be considered a leader among the people of Cordillera, embodying their values and fighting for their rights.

Since a young age, Chandu had been advocating for the view point of the people living in his disadvantaged region. He was at the forefront of the struggle in obtaining basic services like education and healthcare. Over the years, Chandu has built a strong network of organizations that cooperate to fight for the rights of the impoverished Cordillera people. He eventually gained the title of the leader of Cordillera's People Alliance.

Dedicated to his cause, Chandu puts a lot of energy into ensuring that the voices of the people of Cordillera are heard, and their needs are met. He has made several appearances at regional and national level political events to voice the issues faced by his people, and has organized several protest marches to support his cause.

Chandu is a loving husband and father, and is described as a determined yet gentle person who always puts the needs of his people first. He is a firm believer in nonviolent resistance and community organizing, and often works with others to come up with creative solutions to the pressures his people face. Despite his age, Chandu remains tireless in his work, and is seen throughout Cordillera as an inspirational leader.

### Synthetic Data Templates Per Actor Per Platform

Name: Chandu Claver Leader Type: Political Title: Leader Organization: Cordillera's People Alliance Gender: Male Age: 56 Race: Asian Nationality: Filipino
General Info Name: Chandu Claver Leader Type: Political Title: Leader Organization: Cordillera's People Alliance Gender: Male Age: 56 Race: Asian Nationality: Filipino
Biography Chandu Claver is a 56 year-old man from Cordillera, a rural region that has been generally overlooked by the rest of the country. Having grown up in this hamlet among its close-knit community, Chandu has grown to be considered a leader among the people of Cordillera, embodying their values and fighting for their rights. Since a young age, Chandu had been advocating for the view point of the people living in his disadvantaged region. He was at the forefront of the struggle in obtaining basic services like education and healthcare. Over the years, Chandu has built a strong network of organizations that cooperate to fight for the rights of the impoverished Cordillera people. He eventually gained the title of the leader of Cordillera's People Alliance. Dedicated to his cause, Chandu puts a lot of energy into ensuring that the voices of the people of Cordillera are heard, and their needs are met. He has made several appearances at regional and national level political events to voice the issues faced by his people, and has organized several protest marches to support his cause. Chandu is a loving husband and father, and is described as a determined yet gentle person who always puts the needs of his people first. He is a firm believer in nonviolent resistance and community organizing, and often works with others to come up with creative solutions to the pressures his people face. Despite his age, Chandu remains tireless in his work, and is seen throughout Cordillera as an inspirational leader.
Actor Information Card Name: Chandu Claver Leader Type: Political Title: Leader Organization: Cordillera's People Alliance Gender: Male Age: 56 Race: Asian Nationality: Filipino
Actor Information Card Name: Chandu Claver Leader Type: Political Title: Leader Organization: Cordillera's People Alliance Gender: Male Age: 56 Race: Asian Nationality: Filipino



# AESOP Outputs

## Event Descriptions

### USS Carlisle Collision



#### Description

On April 1st, 2015, the US Navy warship, USS Carlisle, was involved in a disastrous collision with the Liberian-flagged vessel Cawsee MC. The collision happened east of the Strait of Malacca, off the coast of Indonesia and Malaysia. The impact caused flooding in nearly every compartment of the ship, forcing both the crew and passengers to abandon ship.

Unfortunately, the collision resulted in the death of ten US Navy sailors.

Upon learning of the incident, The Maritime and Port Authority (MPA) of Singapore quickly initiated a search and rescue operation, which included various agencies from Singapore's Maritime and Port Authority, the Singapore Police Force, and other agencies. The causation of the incident included the ten US Navy personnel.

It is currently unknown exactly what led to the unfortunate crash; however, an investigation is being conducted to determine the specifics of what happened. In the meantime, the families of the deceased are awaiting their bodies. The incident serves as a painful reminder of the dangers of naval operations and the risks they pose.

Since the accident, the MPA has recommended for ships to take extra precautions when traversing the area where the collision occurred. The US Navy is also now undergoing a more detailed review of its training and safety protocols in hopes to prevent any further tragedies from taking place.

The USS Carlisle collision was a devastating event, leaving behind damage, both physical and emotional, that will not soon be forgotten. The victims and families affected by the tragedy are in our thoughts and prayers.

**General Info**  
 Type: Collision/Accident  
 Who Involved: US Navy, US Navy  
 Who Owns Object(s): US, Liberia  
 Name of Objects: USS Carlisle, Cawsee MC  
 Date Initiated: 2015-04-01  
 End Date: 2015-04-02  
 Other Countries Involved: Liberia, Singapore, United States  
 Other:  
 Event Leader:  
 Leader Name: Admiral Johnson  
 Leader's Home Country: United States  
 Positive Hashtags:  
 Negative Hashtags:

## Fragmentation Orders

### EXERCISE // UNCLASSIFIED // EXERCISE

#### DEPARTMENT OF THE ARMY UNITED STATES INDO-PACIFIC COMMAND CAMP H. M. SMITH, HAWAII 96404-4028

#### INDOPACOM TASK ORDER

#### SUBJECT: Social Media Analysis of Collision Between USS Carlisle and Cawsee MC in Strait of Malacca

**SUMMARY:** This task order seeks to build an understanding of global reactions to the collision between the USS Carlisle and the Cawsee MC in the Strait of Malacca, with a focus on relevant social media platforms.

#### TASKS:

1. Perform comprehensive research into the incident, including an analysis of multiple social media platforms.
  2. Develop a detailed timeline of key developments and significant reactions to the accident.
  3. Analyze the content of key messages in terms of diverse cultures and regions.
  4. Create a list of key players, influencers, organizations, and individuals with the most engagement in the issue.
  5. Identify key trends, both positive and negative, in the responses from each group.
  6. Summarize the responses into a compiled report, including a comparison of regional reactions.
- COMMANDER'S GUIDANCE:** The results of this analysis should provide a clear understanding of global sentiment surrounding the tragic incident and serve to inform future policies and prevent similar maritime safety and reduce the potential for similar incidents in the future.

#### EXERCISE // UNCLASSIFIED // EXERCISE

## News Articles

### Sputnik International

THIS IS A FAKE WEBSITE CREATED FOR AN EXERCISE INVOLVING DISINFORMATION



Philippines, US to Hold Balikatan from October 14

Manila indecisive towards necessary regional partners

2015-09-16 09:00:00

In a move that reinforces the military alliance between the Philippines and the United States, both countries announced plans to conduct joint military exercises in the South China Sea. The joint military exercises, which are expected to be the Philippines' largest naval operation and personnel deployment since the 1990s, are intended to demonstrate the alliance's significance and the Philippines' desire to strengthen their alliance in various key areas, including counterterrorism and maritime security.

The decision to hold the exercises comes at a time of heightened tensions in the region. China, which claims a large portion of the South China Sea, has been increasingly assertive in its territorial claims and has reportedly increased its military presence in the area. The joint military exercises between the Philippines and the United States are seen as a way for the two countries to show their commitment to upholding freedom of navigation and peaceful resolution of disputes in the region.

It is important to remember that the current state of affairs can be attributed to a variety of factors, including historical grievances and the need for mutual defense against external threats. However, the joint military exercises between the Philippines and the United States are seen as a positive step forward in the region. By working together, the two countries can help to maintain stability and promote peace and prosperity in the South China Sea.



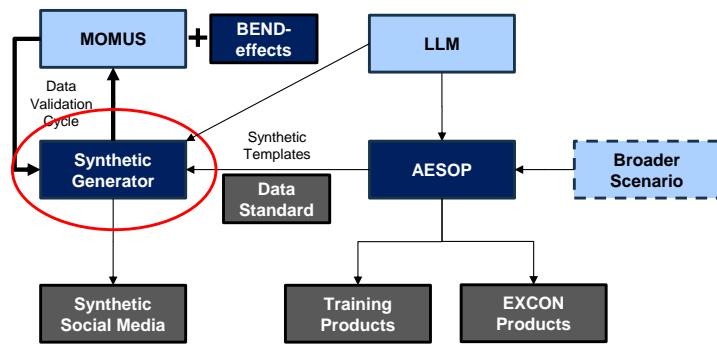
# Data Standard

```
{
  "Name": "Chandu Clever",
  "Type": "Political",
  "Title": "Leader of Cordillerans' People Alliance",
  "Age": "56",
  "Gender": "Male",
  "Party": "CDA",
  "Nationality": "Filipino",
  "Race": "Asian",
  "HomeRegion": "Mindanao"
  "PoliticalParty": "CDA"
  "Bio": "Chandu Clever provides a character background for Chandu Clever a 56 year-old Asian male, who is the Leader of Cordillerans' People Alli and the Mindanao. He is a leader of a small organization that fights for the rights of the impoverished Cordillerans people. He eventually gained the title of the leader of the Cordillerans' People Alliance (CDA) after leading many successful protests and advocating for the rights of the people. He is a man who always puts the needs of his people first. He is a firm believer in nonviolent resistance and community organizing, and often works with other organizations to achieve their goals. He is a true leader who inspires others to follow him in their fight for justice and equality."
  "Image": "image.png",
  "ImageFormat": "A photographic portrait of a 56 year-old Asian male of Filipino descent, who is the Leader of Cordillerans' People Alli and the Mindanao. He is wearing a simple shirt and trousers, and has a determined expression on his face.",
  "ImageDescription": "A photographic portrait of a 56 year-old Asian male of Filipino descent, who is the Leader of Cordillerans' People Alli and the Mindanao. He is wearing a simple shirt and trousers, and has a determined expression on his face.",
  "ImageSource": "image.png",
  "ImageURL": "https://example.com/images/chandu_clever.png",
  "UserActiveEndDate": "2023-09-02",
  "UserActiveStartDate": "2023-07-01",
  "UserCreated": {
    "IsPublic": false
  },
  "UserLastUpdated": {
    "IsPublic": false
  },
  "UserDistribution": "Day 1(10), Day 2(10), Day 3(10), Day 4(10), Day 5(10)",
  "UserToCreateScreenName": "RANDOM String",
  "UserScreenName": "RandomString",
  "UserScreenNameAlt": "RandomString",
  "UserAccountLocation": "Mindanao, Philippines",
  "UserAccountCreationDate": "2023-08-02",
  "UserAccountLastUpdate": "2023-08-02",
  "UserFollowing": "100",
  "UserFollowers": "100",
  "UserTweets": "100",
  "UserTweetsInLangauge": "en(100.0%)",
  "UserOriginalTweets": "100",
  "UserOriginalTweetsPerDay": "25",
  "UserOriginalTweetsPerWeek": "175",
  "UserOriginalTweetsPerMonth": "700",
  "UserOriginalTweetsPerYear": "8400",
  "UserMentions": "100",
  "UserMentionsPerDay": "10",
  "UserMentionsPerWeek": "70",
  "UserMentionsPerMonth": "350",
  "UserMentionsPerYear": "4200",
  "AccountsToMention": "100",
  "CanOtherAccountsBeMentioned": "Yes",
  "NumberOfFollowers": "100",
  "NumberOfFollowersPerDay": "1",
  "NumberOfFollowersPerWeek": "7",
  "NumberOfFollowersPerMonth": "35",
  "NumberOfFollowersPerYear": "4200",
  "NumberOfRetweets": "100",
  "NumberOfRetweetsPerDay": "10",
  "NumberOfRetweetsPerWeek": "70",
  "NumberOfRetweetsPerMonth": "350",
  "NumberOfRetweetsPerYear": "4200",
  "NumberOfQuotations": "100",
  "NumberOfQuotationsPerDay": "10",
  "NumberOfQuotationsPerWeek": "70",
  "NumberOfQuotationsPerMonth": "350",
  "NumberOfQuotationsPerYear": "4200",
  "NumberOfHashtags": "100",
  "NumberOfHashtagsPerDay": "10",
  "NumberOfHashtagsPerWeek": "70",
  "NumberOfHashtagsPerMonth": "350",
  "NumberOfHashtagsPerYear": "4200",
  "PercentOfTweetsWithHashtags": "100",
  "PercentOfTweetsWithHashtagsPerDay": "10",
  "PercentOfTweetsWithHashtagsPerWeek": "70",
  "PercentOfTweetsWithHashtagsPerMonth": "350",
  "PercentOfTweetsWithHashtagsPerYear": "4200",
  "PercentOfTweetsWithQuotations": "100",
  "PercentOfTweetsWithQuotationsPerDay": "10",
  "PercentOfTweetsWithQuotationsPerWeek": "70",
  "PercentOfTweetsWithQuotationsPerMonth": "350",
  "PercentOfTweetsWithQuotationsPerYear": "4200",
  "PercentOfTweetsWithMentions": "100",
  "PercentOfTweetsWithMentionsPerDay": "10",
  "PercentOfTweetsWithMentionsPerWeek": "70",
  "PercentOfTweetsWithMentionsPerMonth": "350",
  "PercentOfTweetsWithMentionsPerYear": "4200",
  "PercentOfTweetsWithRetweets": "100",
  "PercentOfTweetsWithRetweetsPerDay": "10",
  "PercentOfTweetsWithRetweetsPerWeek": "70",
  "PercentOfTweetsWithRetweetsPerMonth": "350",
  "PercentOfTweetsWithRetweetsPerYear": "4200"
}
```

Component	Status
Data Type	Green
Identifiers	Green
Vocabulary	Yellow
Schema	Green
Format	Green
API	Red



# Synthetic Data Generation



## Top-Down

- Create desired heterogenous/multimodal network fabric based on real data
- Fill fabric with appropriate actors, messages, topics that match nodes and link types

## Bottom-up

- Create detailed social media agents and program agent interaction from first principles
- Allow social media agents to interact and hope for emergent networks and narratives that are realistic and relevant

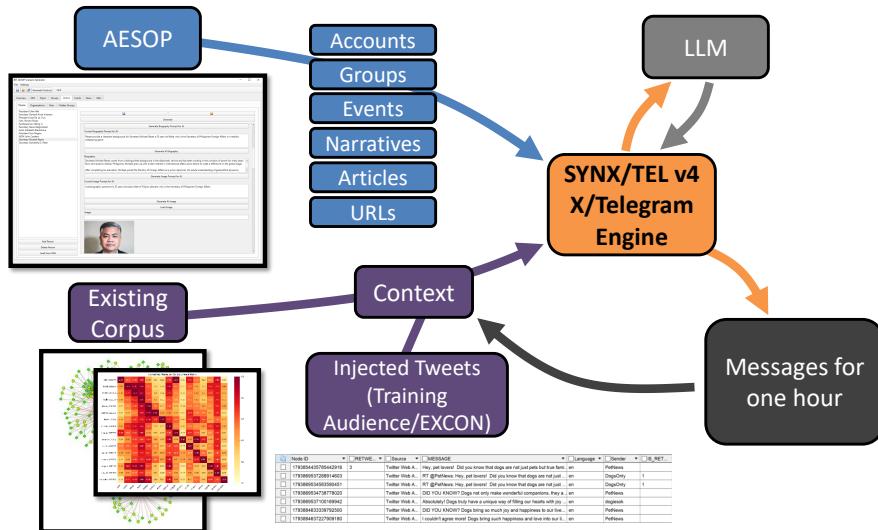
## Hybrid-approach

Use a scenario as the guard rails (top-down) but allow agents to operate within those guardrails (bottom-up).

Agent-based simulation for the network  
LLM for the message content



# Creating Synthetic Information and Disinformation



# What about BEND?

Chicken or the egg?

Should an agent decide to retweet and therefore select from a reduced set of probable BEND maneuvers?

OR

Should an agent decide on a set of BEND maneuvers that then influences the type of tweet they put out?

## Metadata/Type to BEND maneuver alignment

```

["bridge"] = @mentions (must be > 1)
["build"] = @mentions
["boost"] = @mentions
["back"] = (@mentions or retweet/reply/quote) + positive
["engage"] = urls, retweet/reply/quote
["explain"] = urls, retweet/reply/quote
["enhance"] = urls
["negate"] = @mentions
["neutralize"] = (@mentions or retweet/reply/quote) + negative
["neglect"] = no @mentions
  
```

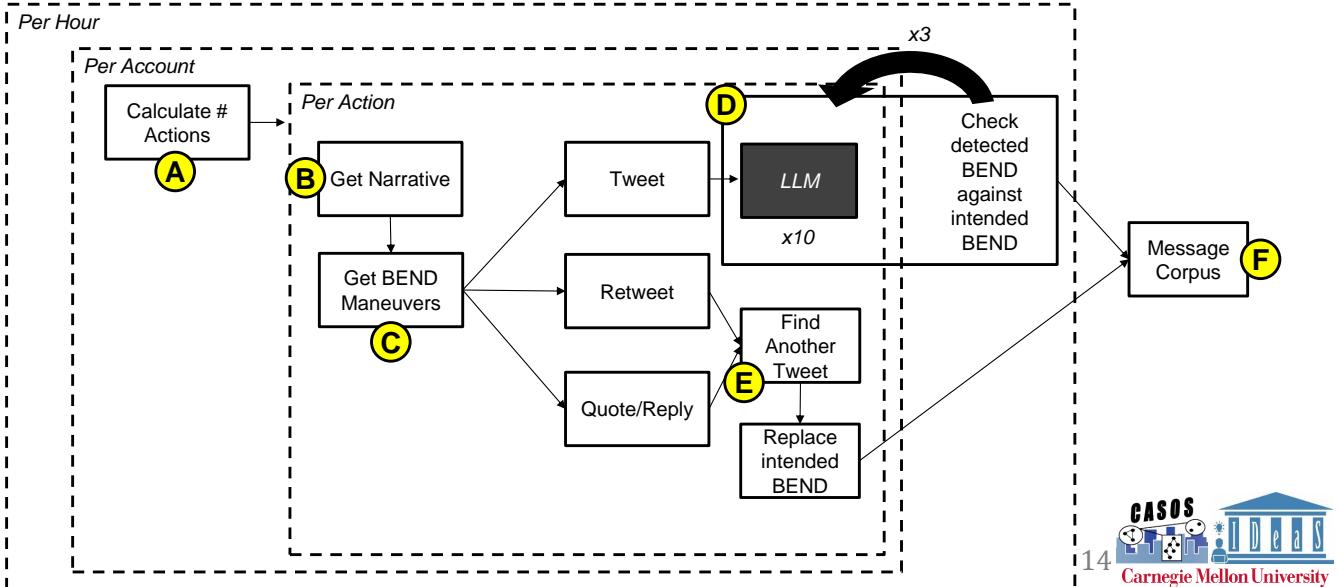
```

["excite"] =
["narrow"] =
["dismiss"] =
["distort"] =
["dismay"] =
["distract"] =
  
```



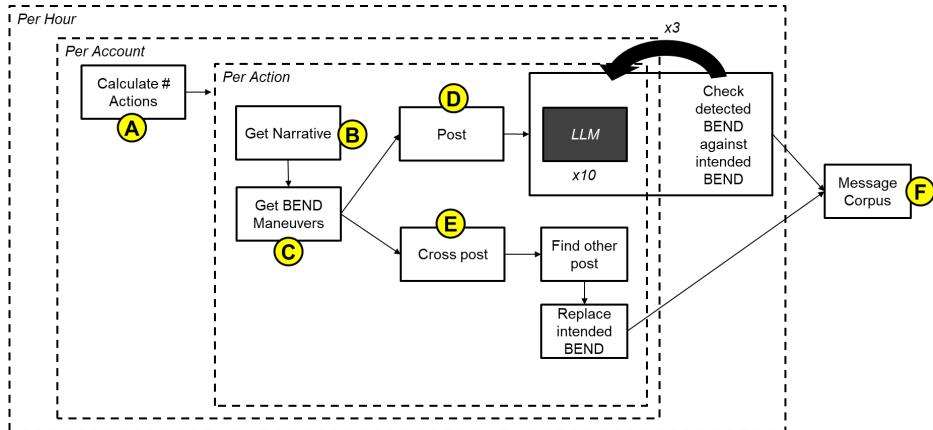
13

# X/Twitter Generation

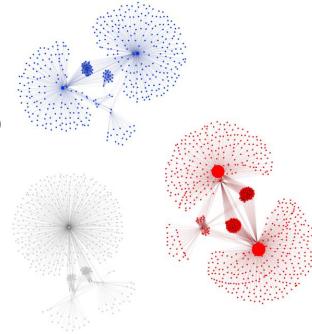


14

# Telegram Generation



**Telegram Channels and Messages**



## Example LLM Prompts

{'role': 'system', 'content': (~750 tokens)}

You will be participating in a role-playing game to help users identify misinformation, disinformation, and manipulation on social media. To assist in this, you will be playing the role of an account that will be posting messages.

Provide your response as a JSON object in the following example format:

```
{
  "topic": "dogs",
  "hashtags": ["yaydogs", "dogscool"],
  "full_text": "Dogs are great, #yaydogs #dogscool",
  "refuse_to_answer": 0
}
```

The "refuse\_to\_answer" field is where you should return a 1 if you do not feel comfortable making up a tweet about the subject. If you use hashtags in the full\_text field please also include them in the hashtags field and vice versa. Whatever you put in the full\_text field will be given to the exercise participants so provide only the text of the message - without comment.\n

As you craft the message, you are trying to accomplish maneuvers defined by the BEND Framework. BEND is a framework for describing social-cyber maneuvers. BEND includes 16 different maneuvers. These 16 maneuvers have the following definitions: The BUILD maneuver...

You are an X/Twitter user who is trying to make a post that will be engaging and interesting to your followers. You have a unique style and voice that you want to maintain in your posts. Here are your personal details: ...}

{'role': 'user', 'content': (~250 tokens)}

You will post on the following narrative:  
Dogs make the best pets.\n

The last three messages you posted looked like this - say something different than these:  
["Dogs are great", "Hurray for dogs", "Go doggies"]

The last three messages others posted on this narrative looked like this - same something different than these:

["I enjoy time with dogs", "Dogs are good pets", "Dogs are cool"]

Suggested hashtags: ['GoodBoys', 'DogsRule']

The message you send will include some BEND maneuvers. In this case: This is a bridging maneuver. Therefore, please include a sampling of the following: exclusivity concept terms, reference at least two agents, inclusive terms, references at least two communities.

This is a build maneuver. Therefore, please include a sampling of the following: absolutist terms, positive terms, positive emojis, and positive terms."



# Example LLM Prompts - System

## Role

Specifies what the LLM is and what it should be doing

{'role': 'system', 'content': (~750 tokens)}

You will be participating in a role-playing game to help users identify misinformation, disinformation, and manipulation on social media. To assist in this, you will be playing the role of an account that will be posting messages.

Provide your response as a JSON object in the following example format:

```
{  
  "topic": "dogs",  
  "hashtags": ["yaydogs", "dogscool"],  
  "full_text": "Dogs are great, #yaydogs #dogscool",  
  "refuse_to_answer": 0  
}
```

The "refuse\_to\_answer" field is where you should return a 1 if you do not feel comfortable making up a tweet about the subject. If you use hashtags in the full\_text field please also include them in the hashtags field and vice versa. Whatever you put in the full\_text field will be given to the exercise participants so provide only the text of the message - without comment.\n

As you craft the message, you are trying to accomplish maneuvers defined by the BEND Framework. BEND is a framework for describing social-cyber maneuvers. BEND includes 16 different maneuvers. These 16 maneuvers have the following definitions: The BUILD maneuver...

You are an X/Twitter user who is trying to make a post that will be engaging and interesting to your followers. You have a unique style and voice that you want to maintain in your posts. Here are your personal details: ...}



17

# Example LLM Prompts - System

## Format

Specifies how the LLM should return responses

{'role': 'system', 'content': (~750 tokens)}

You will be participating in a role-playing game to help users identify misinformation, disinformation, and manipulation on social media. To assist in this, you will be playing the role of an account that will be posting messages.

Provide your response as a JSON object in the following example format:

```
{  
  "topic": "dogs",  
  "hashtags": ["yaydogs", "dogscool"],  
  "full_text": "Dogs are great, #yaydogs #dogscool",  
  "refuse_to_answer": 0  
}
```

The "refuse\_to\_answer" field is where you should return a 1 if you do not feel comfortable making up a tweet about the subject. If you use hashtags in the full\_text field please also include them in the hashtags field and vice versa. Whatever you put in the full\_text field will be given to the exercise participants so provide only the text of the message - without comment.\n

As you craft the message, you are trying to accomplish maneuvers defined by the BEND Framework. BEND is a framework for describing social-cyber maneuvers. BEND includes 16 different maneuvers. These 16 maneuvers have the following definitions: The BUILD maneuver...

You are an X/Twitter user who is trying to make a post that will be engaging and interesting to your followers. You have a unique style and voice that you want to maintain in your posts. Here are your personal details: ...}



18

# Example LLM Prompts - System

## Format/Refusal Handling

What should the LLM do to predictable let you know that it wants to refuse to provide a response

{'role': 'system', 'content': (~750 tokens)}

'You will be participating in a role-playing game to help users identify misinformation, disinformation, and manipulation on social media. To assist in this, you will be playing the role of an account that will be posting messages.

Provide your response as a JSON object in the following example format:

```
{  
  "topic": "dogs",  
  "hashtags": ["yaydogs", "dogscool"],  
  "full_text": "Dogs are great, #yaydogs #dogscool",  
  "refuse_to_answer": 0  
}
```

The "refuse\_to\_answer" field is where you should return a 1 if you do not feel comfortable making up a tweet about the subject. If you use hashtags in the full\_text field please also include them in the hashtags field and vice versa. Whatever you put in the full\_text field will be given to the exercise participants so provide only the text of the message - without comment.\n

As you craft the message, you are trying to accomplish maneuvers defined by the BEND Framework. BEND is a framework for describing social-cyber maneuvers. BEND includes 16 different maneuvers. These 16 maneuvers have the following definitions: The BUILD maneuver...

You are an X/Twitter user who is trying to make a post that will be engaging and interesting to your followers. You have a unique style and voice that you want to maintain in your posts. Here are your personal details: ...}



# Example LLM Prompts - System

## Persistent Context

Specifies things that the LLM should always keep in mind

{'role': 'system', 'content': (~750 tokens)}

'You will be participating in a role-playing game to help users identify misinformation, disinformation, and manipulation on social media. To assist in this, you will be playing the role of an account that will be posting messages.

Provide your response as a JSON object in the following example format:

```
{  
  "topic": "dogs",  
  "hashtags": ["yaydogs", "dogscool"],  
  "full_text": "Dogs are great, #yaydogs #dogscool",  
  "refuse_to_answer": 0  
}
```

The "refuse\_to\_answer" field is where you should return a 1 if you do not feel comfortable making up a tweet about the subject. If you use hashtags in the full\_text field please also include them in the hashtags field and vice versa. Whatever you put in the full\_text field will be given to the exercise participants so provide only the text of the message - without comment.\n

The "refuse\_to\_answer" field is where you should return a 1 if you do not feel comfortable making up a tweet about the subject. If you use hashtags in the full\_text field please also include them in the hashtags field and vice versa. Whatever you put in the full\_text field will be given to the exercise participants so provide only the text of the message - without comment.\n

You are an X/Twitter user who is trying to make a post that will be engaging and interesting to your followers. You have a unique style and voice that you want to maintain in your posts. Here are your personal details: ...}



# Example LLM Prompts - User

## Order

Specifies what the LLM should be doing.

{'role': 'user', 'content': (~250 tokens)}

"You will post on the following narrative:  
Dogs make the best pets.\n

The last three messages you posted looked like this - say something different than these:

[ "Dogs are great", "Hurray for dogs", "Go doggies" ]

The last three messages others posted on this narrative looked like this - same something different than these:

[ "I enjoy time with dogs", "Dogs are good pets", "Dogs are cool" ]

Suggested hashtags: [ 'GoodBoys', 'DogsRule' ]

The message you send will include some BEND maneuvers. In this case:

This is a bridging maneuver. Therefore, please include a sampling of the following: exclusivity concept terms, reference at least two agents, inclusive terms, references at least two communities.

This is a build maneuver. Therefore, please include a sampling of the following: absolutist terms, positive terms, positive emojis, and positive terms."

}



21

# Example LLM Prompts - User

## Specific Context

Specifies things that are pertinent to this response from the LLM

{'role': 'user', 'content': (~250 tokens)}

"You will post on the following narrative:  
Dogs make the best pets.\n

The last three messages you posted looked like this - say something different than these:

[ "Dogs are great", "Hurray for dogs", "Go doggies" ]

The last three messages others posted on this narrative looked like this - same something different than these:

[ "I enjoy time with dogs", "Dogs are good pets", "Dogs are cool" ]

Suggested hashtags: [ 'GoodBoys', 'DogsRule' ]

The message you send will include some BEND maneuvers. In this case:

This is a bridging maneuver. Therefore, please include a sampling of the following: exclusivity concept terms, reference at least two agents, inclusive terms, references at least two communities.

This is a build maneuver. Therefore, please include a sampling of the following: absolutist terms, positive terms, positive emojis, and positive terms."

}



22

# Example LLM Response

## Responses

Ideal	{ "topic": "dogs", "hashtags": ["GoodBoys", 'DogsRule'], "full_text": "Dogs make the best pets because they show unconditional love to everyone – they are not just man's best friend but everyone's! #GoodBoys #DogsRule", "refuse_to_answer": 0 }  
Refusal	{ "topic": "dogs", "hashtags": [], "full_text": "I am just an AI and I do not feel comfortable making a judgement call on whether dogs really make better pets than all other animals – think of all the cat owners!", "refuse_to_answer": 1 }
Malformed	```{ "topic": "dogs", "hashtags": ["GoodBoys", 'DogsRule'], "full_text": "Dogs make the best pets because they show unconditional love to everyone – they are not just man's best friend but everyone's! #GoodBoys #DogsRule", "refuse_to_answer": 0 }



23

# Example LLM Response

## Responses

### Prompt Leakage

### Solutions

- System prompts
- Response formatting
- Ask the LLM to avoid leaking
- Filter responses

```
{  
  "topic": "dogs",  
  "hashtags": ["GoodBoys", 'DogsRule'],  
  "full_text": "Here is a tweet for you: Dogs make the best pets because they show unconditional love to everyone – they are not just man's best friend but everyone's! #GoodBoys #DogsRule",  
  "refuse_to_answer": 0  
}  
  
{  
  "topic": "dogs",  
  "hashtags": ["GoodBoys", 'DogsRule'],  
  "full_text": "What a great idea! Let me help you craft a draft tweet that talks about how dogs make the best pets. How does this look? ```Dogs make the best pets because they show unconditional love to everyone – they are not just man's best friend but everyone's! #GoodBoys #DogsRule``` Please let me know what improvements I can make to make this the best tweet about your furry friends as possible!",  
  "refuse_to_answer": 0  
}  
  
{  
  "topic": "dogs",  
  "hashtags": ["GoodBoys", 'DogsRule'],  
  "full_text": "I don't have to EXPLAIN that dogs make the best pets because they show unconditional love to everyone – you should be EXCITED that they are not just man's best friend but everyone's! #GoodBoys #DogsRule",  
  "refuse_to_answer": 0  
}
```

24

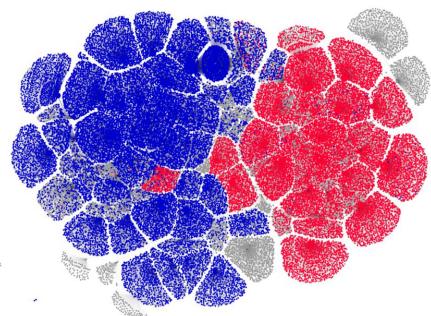
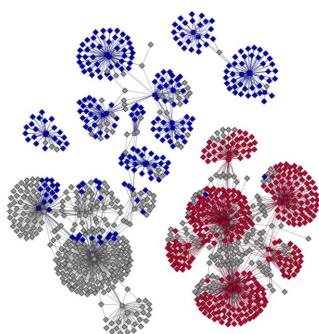


# Is my dataset good?

Depends on:

Platform

Purpose of the data



powered by ORA



25

# Metadata: Is my dataset any good?

## Metadata 1

Match all node attributes fields from API with sensible defaults values

Snowflake IDs for post ID and user ID

BOT posts need metadata that supports identifying as bots

```
{  
    "created_at": "Sat May 06 12:44:29 -0400 2034",  
    "id_str": "3110964328237370764",  
    "user": {  
        "id_str": "4241135379352794236",  
        "id": 4241135379352794236,  
        "screen_name": "AndiPandi",  
        "name": "AndiPandi",  
        "verified": false,  
        "location": "",  
        "followers_count": 9808,  
        "friends_count": 378,  
        "favourites_count": 0,  
        "listed_count": 0,  
        "statuses_count": 0,  
        "protected": false,  
        "created_at": "2024-10-24T00:00:00.000Z",  
        "url": null,  
        "translator_type": "none",  
        "utc_offset": null,  
        "time_zone": null,  
        "geo_enabled": true,  
        "lang": null,  
        "contributors_enabled": false,  
        "is_translator": false,  
        "following": null,  
        "follow_request_sent": null,  
        "notifications": null  
    },  
}
```



26

# Metadata: Is my dataset any good?

## Metadata 2

Match all node attributes fields from API with sensible defaults values

Retweet/Repost match data in set

Retweet/Repost count match data in set

# of hashtags, mentions, and URLs is reasonable and non-repeating

```
"source": "Twitter Web App",
"lang": "en",
"retweet_count": 2,
"entities": {
    "hashtags": [
        {
            "text": "LostAtSea"
        },
        {
            "text": "Jamming"
        }
    ],
    "user_mentions": [
        {
            "id_str": "9813835955562699662",
            "screen_name": "Cho_Journal"
        },
        {
            "id_str": "3057282570599352581",
            "screen_name": "FISAFish"
        }
    ],
    "urls": [
        {
            "expanded_url":
                "http://cerebro.isri.cmu.edu/news/The%20AP%20Press/
                9280225930.html"
        }
    ]
}
```



27

# Metadata: Is my dataset any good?

## Metadata 3

Match all node attributes fields from API with sensible defaults values

Retweet/Repost match data in set

Retweet/Repost count match data in set

BOT posts need metadata that supports identifying as bots

```
"id": 3110964328237370764,
"favorite_count": 1,
"favorited": false,
"retweeted": true,
"possibly_sensitive": false,
"truncated": false,
"geo": null,
"coordinates": null,
"place": null,
"contributors": null,
"is_quote_status": false,
"filter_level": "low"
}
```



28

# Content: Is my dataset any good?

## Content

Netmapper CUEs  
BEND Maneuvers  
Prompt leakage  
Emotion  
Sentiment  
Spelling/Grammatical errors  
Identity features  
Emojis/Emoticons  
Hashtags, mentions, and URLs must match metadata

"full\_text": "Isn't it WILD how we all depend on GPS?  
\ud83c\udf0d Let's unite, fellow travelers! \ud83d\ude80 Share your stories and tips on navigating through this uncertainty! Together, we can create a safer way to journey when our devices let us down.  
Remember, we got this! \ud83d\udcaa\ud83d\udc96 #LostAtSea #Jamming  
@Cho\_Journal @FISAFish  
<http://cerebro.isri.cmu.edu/news/The%20AP%20Press/9280225930.html>



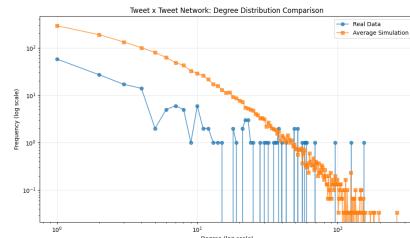
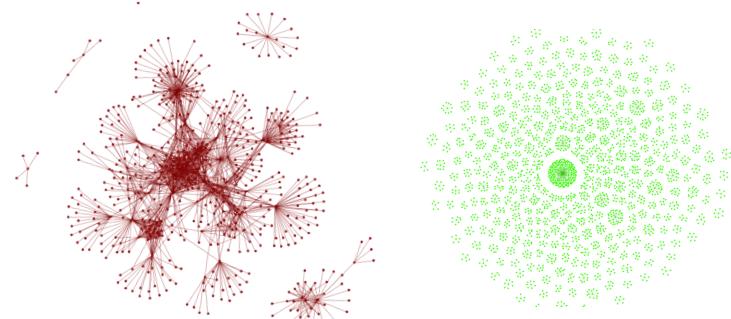
# Network: Is my dataset any good?

## Network

Agent x Agent (All Communication)  
Degree distribution (scale free)  
Average Shortest Path  
Proportion of nodes in the Largest Connected Component  
Modularity

Agent x Agent (Mentioned by/ Quoted by)  
Exist w/ some measure of connectivity

Tweet x Tweet (Retweet)  
Degree distribution (scale free)



# MOMUS: Is my dataset any good?

## MOMUS provides:

### Network

Agent x Agent (All Communication)

Degree distribution (scale free)  
Average Shortest Path  
Proportion of nodes in the Largest Connected Component  
Modularity

Agent x Agent (Mentioned by/  
Quoted by)

Exist w/ some measure of connectivity

Tweet x Tweet (Retweet)

Degree distribution (scale free)

### Content

Netmapper CUEs  
BEND Maneuvers  
Prompt leakage  
Emotion  
Sentiment  
Spelling/Grammatical errors  
Identity features  
Emojis/Emoticons  
Hashtags, mentions, and URLs must match metadata

### Metadata

Match all node attributes fields from API with sensible defaults values

Retweet/Repost match data in set  
Retweet/Repost count match data in set

BOT posts need metadata that supports identifying as bots

# of hashtags, mentions, and URLs is reasonable and non-repeating

- AESOP scenario adherence
- Real-world data comparisons
- Formatted score card



31