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.

The research for this thesis was supported in part by the Office of Naval Research (ONR) under grant N00014182106, the United States Army under grant W911NF20D0002, and the center for Informed Democracy and Social-cybersecurity (IDeaS). 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 ONR, the United States 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.

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

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.[11] 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.[15] [14] I will apply BEND maneuver detection again, both through the current methodology[14] 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
COVID-19 Vaccines (During Rollout)	~2020-12-07 thru 2020-12-14	1,648,309	848,490
Ukraine-Russia Conflict	~2021-11-01 thru 2022-11-06	4,529,740	1,674,753

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".

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 gathered two weeks' worth of Twitter data spanning from October 28 to November 4, 2020, covering the event.

COVID-19 Vaccine This dataset is comprised of tweets discussing COVID-19 and the Pfizer vaccine. Theses tweets were collected using pandemic-related keywords and further refined to focus on vaccine discourse. They were collected from three distinct timeframes surrounding the introduction of the Pfizer/BioNTech vaccine: December 1-7, 2020 (preceding the rollout), December 8-10, 2020 (coinciding with the vaccine's deployment in the US and UK), and January

25-31, 2021 (six weeks post-rollout). These are augmented by additional segments collated by Janice Blaine[14], specifically addressing conspiracy theories and vaccine-related discussions. In total, this dataset captures a year-long narrative of the pandemic's impact.

Ukraine-Russia 2022 This dataset consists of Twitter data sourced sourced from November 2021 through November 2022. It encompasses a wide array of key terms in Russian, Ukrainian, and English, focusing on political figures, geographical locations, and other pertinent topics related to the conflict.

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.[23]. 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.[23] 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.[7] 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.[65] 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 [27]. Cyberspace simulations have been used to model these conflicts, often closely emulating current physical military doctrine [36]. However, these simulations focus primarily on the cyber-terrain itself - accurately deducing that terrain has a large impact on the outcome of conflict [36]. However, just as the physical domain of warfare stretches into the digital space, so too does the information environment. This is social cyber security [22] [21]. Social cyber security welds the methodologies of the social sciences with the need to identify, assess, and counter the impact of information maneuvers.[21]

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[30] and within services there is Branch doctrine[32] - 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.[57] Also, the ABC framework developed by Camille Francois[38], which looks at the Actors, Behaviors, and Content of a disinformation campaign and its successors ABCD[9] and ABCDE.[61] Finally, the SCOTCH framework - focusing on Sources, Channels, Objectives, Targets, Composition, and Hooks - was brought forward by Blazek in 2021.[17]

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 [13]. 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 [13] as refined and validated by Blane et al. in 2022 [15] and later in Blane's thesis

work.[14] 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.[14]

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 [14] is the cutting edge in BEND maneuver detection. This method has seen several iterative improvements - first in Uyheng et al. in 2020[66], then by Blane et al. in 2022[15], then Alieva, et al. in 2022[10], before Blane adopted the the current method in her thesis work.

In this method, linguistic cues are extracted from message text using NetMapper software.[23] These NetMapper cues are a proprietary blend of concepts that represent a message's sentiment and the author's emotional state.[21] These particular cues - now referred to as CUES - are then loaded into ORA-Pro[23] - 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

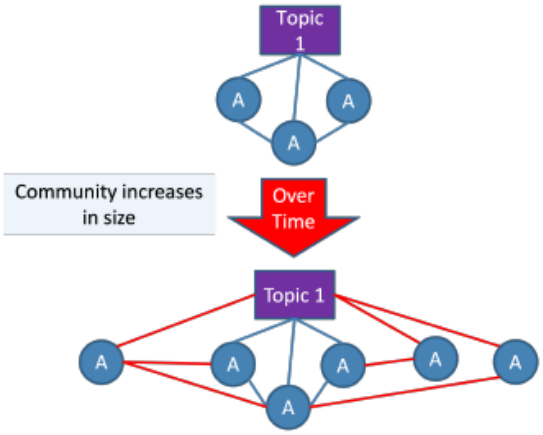

Figure 3.1: Excerpt from Blane's thesis work "Social-Cyber Maneuvers for Analyzing Online Influence Operations".[14]

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[34]
- ADP 3-13 Information[33]
- AFDP 3-13 Information in Air Operation[30]
- CJCSI 3210.01C Joint Information Operations Proponent[24]
- DODI 3600.01 Information Operations[29]
- ADP 5-0 The Operations Process[31]
- MCWP 3-32 Marine Air-Ground Task Force Information Operations[41]
- JP 2-01.3 Joint Intelligence Preparation of the Operational Environment[44]
- ATP 2-01.3 Intelligence Preparation of the Battlefield[32]
- 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 a 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[8].

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)[63]. 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[8], 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 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 [28]. 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

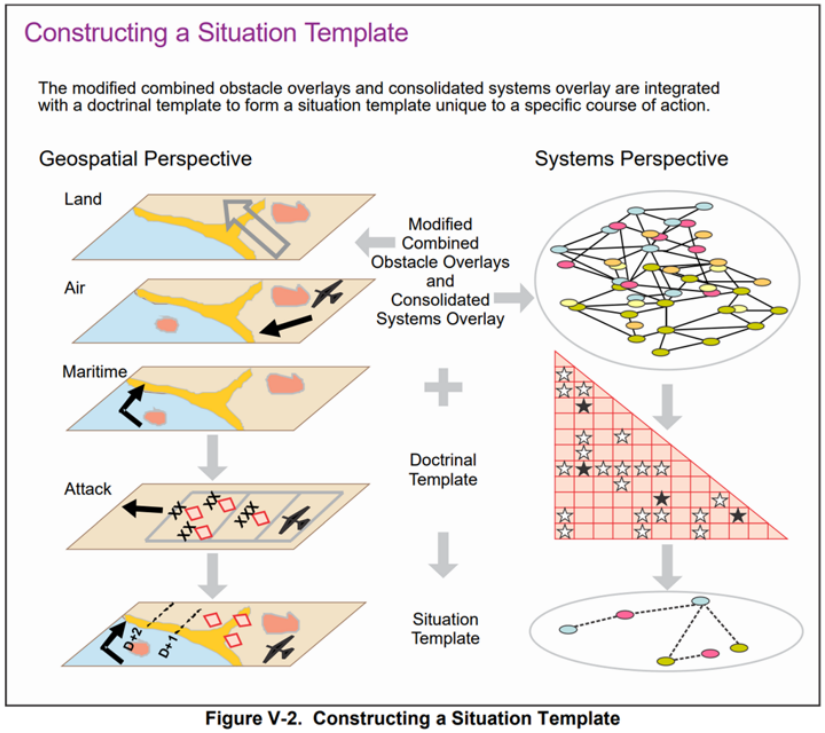

Figure V-2. Constructing a Situation Template

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 A.

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

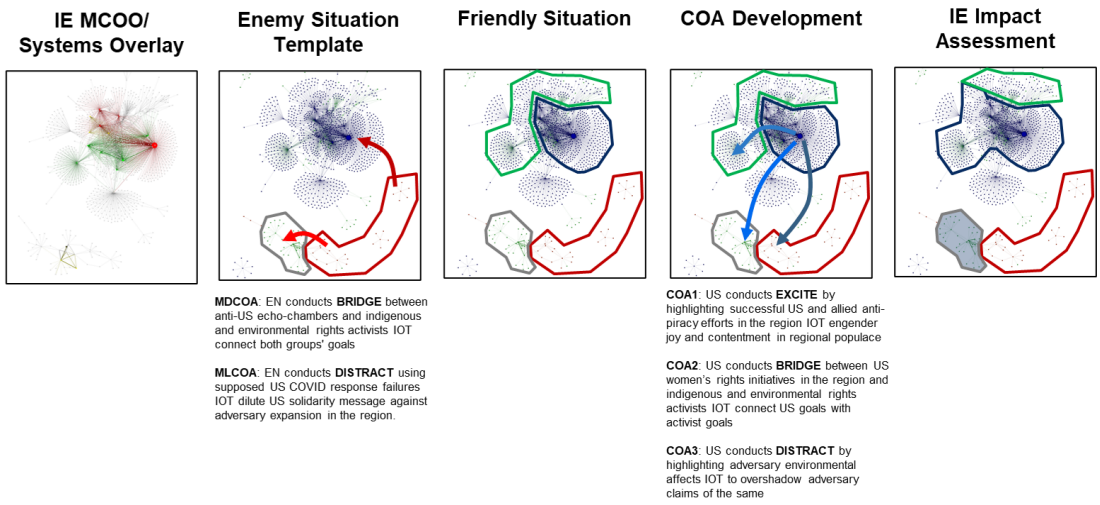

Figure 4.2: Information Environment products influence by BEND as a part of this thesis.

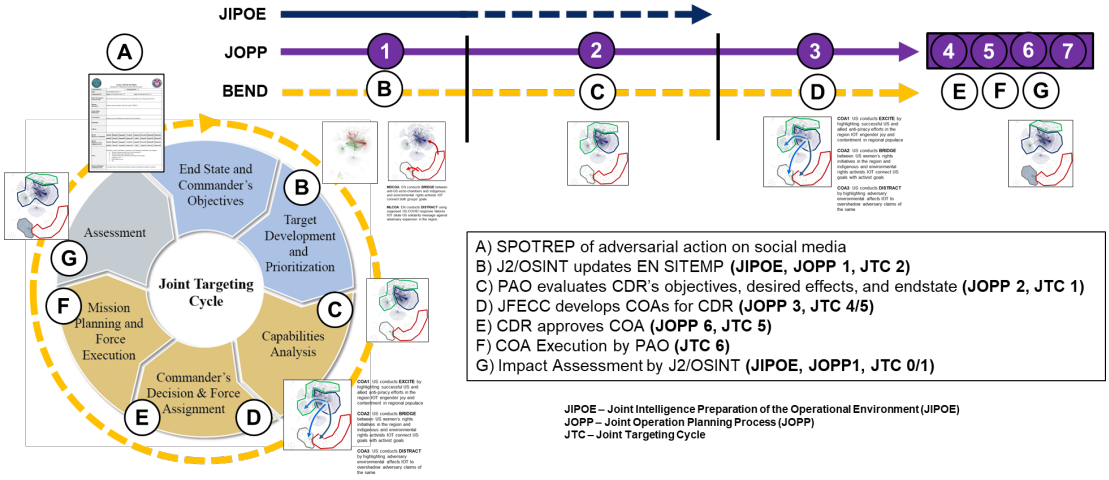

Figure 4.3: BEND integration with the Joint Targeting Cycle, JIPOE, JOPP.

authorities for conducting social media maneuve. 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.[39] 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 is scalable - just collect more - and the gold standard for realism from 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 scalibility - 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.

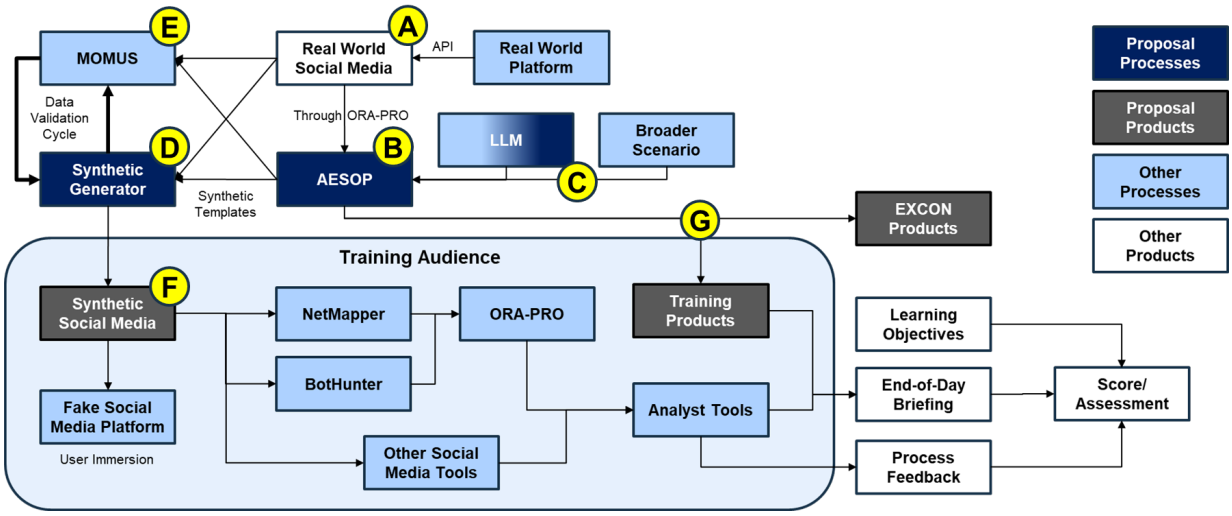

Figure 5.1: Project OMEN training flow.

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.[65] 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.

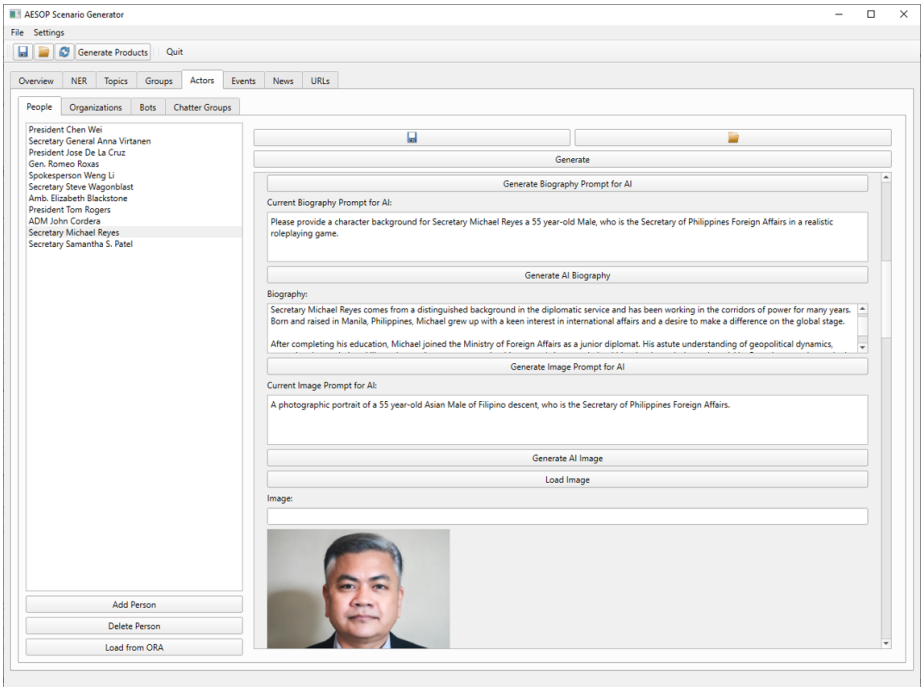

Figure 5.2: AESOP GUI Example. With an actor template being worked on.

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[7] 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.[60]

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.

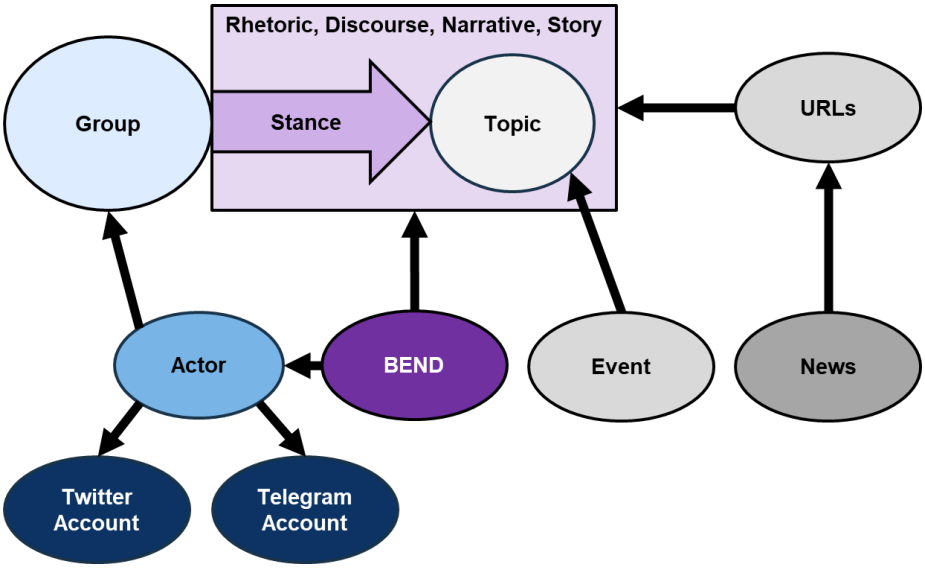

Figure 5.3: AESOP Relational Diagram. Blue is loosely defined as the WHO, purple is HOW, and gray is the WHAT.

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 fro each from -5 to 5 approximating a Likert scale of approval for each topic.

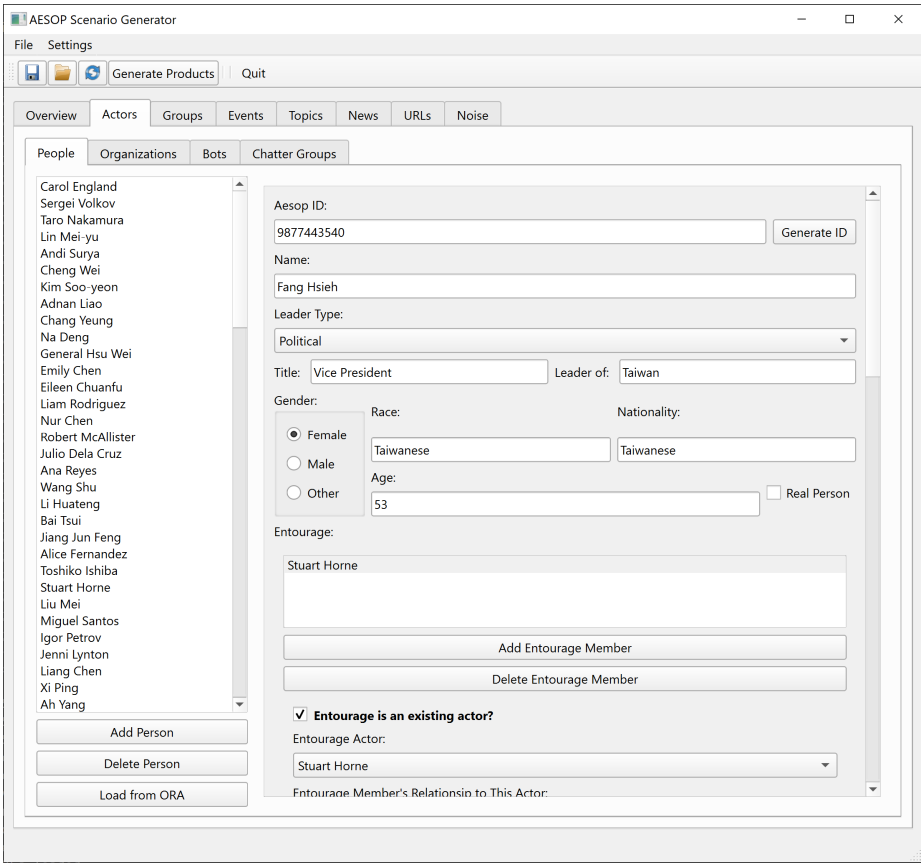

Figure 5.4: AESOP GUI for Actor creation.

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.[69] 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.[18] 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.

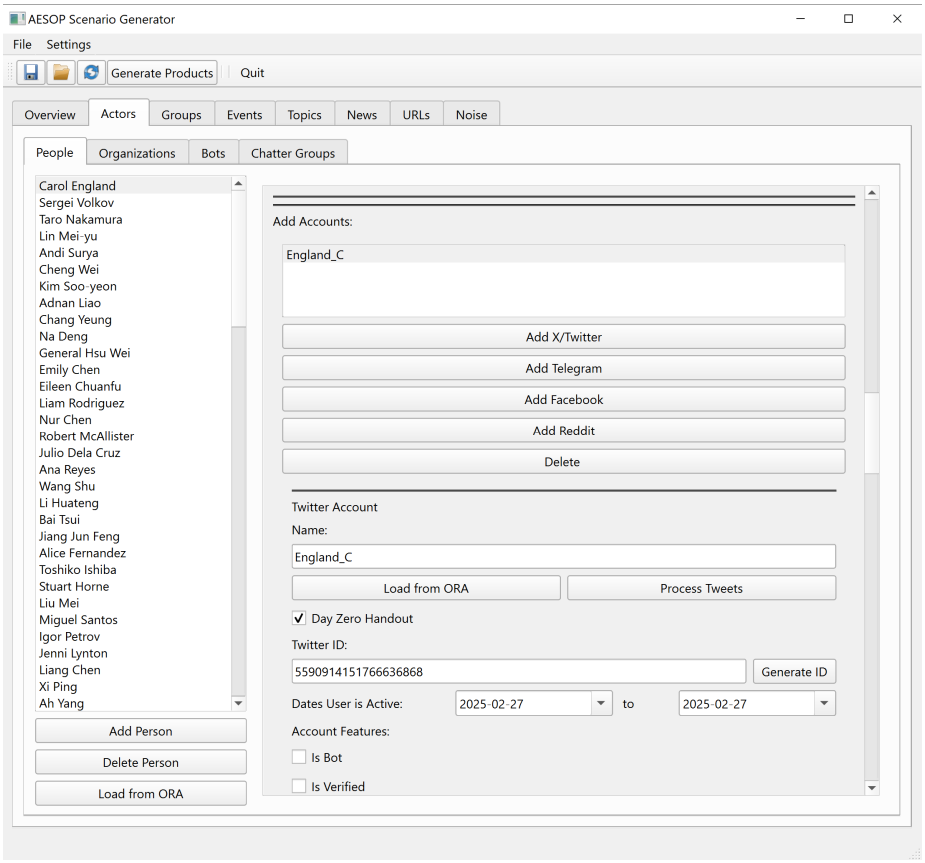

Figure 5.5: AESOP GUI for Account creation.

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.

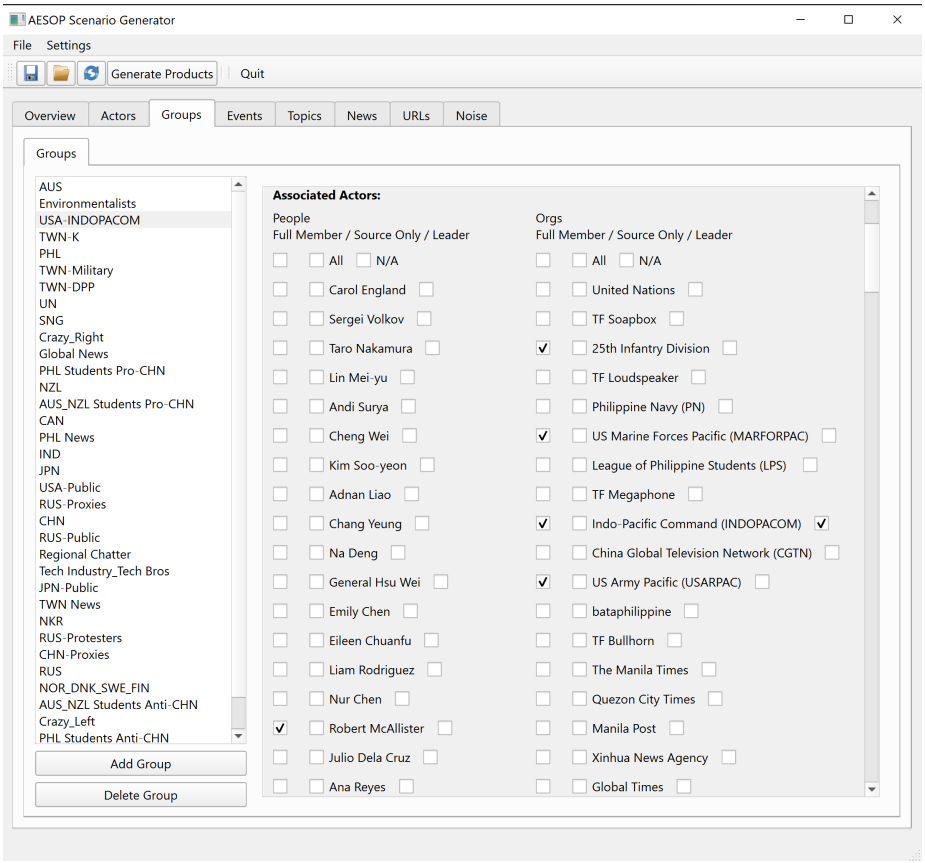

Figure 5.6: AESOP GUI for Group creation.

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
- Postive 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, Fragmentory 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 Fragmentory 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

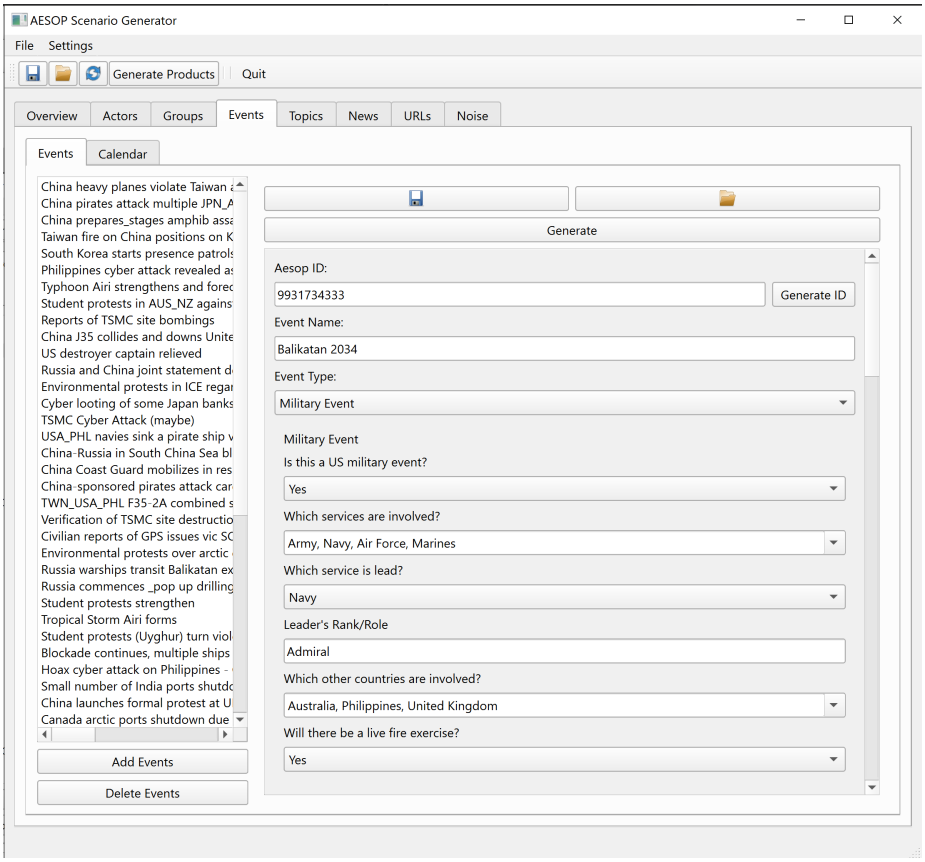

Figure 5.7: AESOP GUI for Event creation.

way - they are exactly what they sound like - a Press Release from a specifed 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.

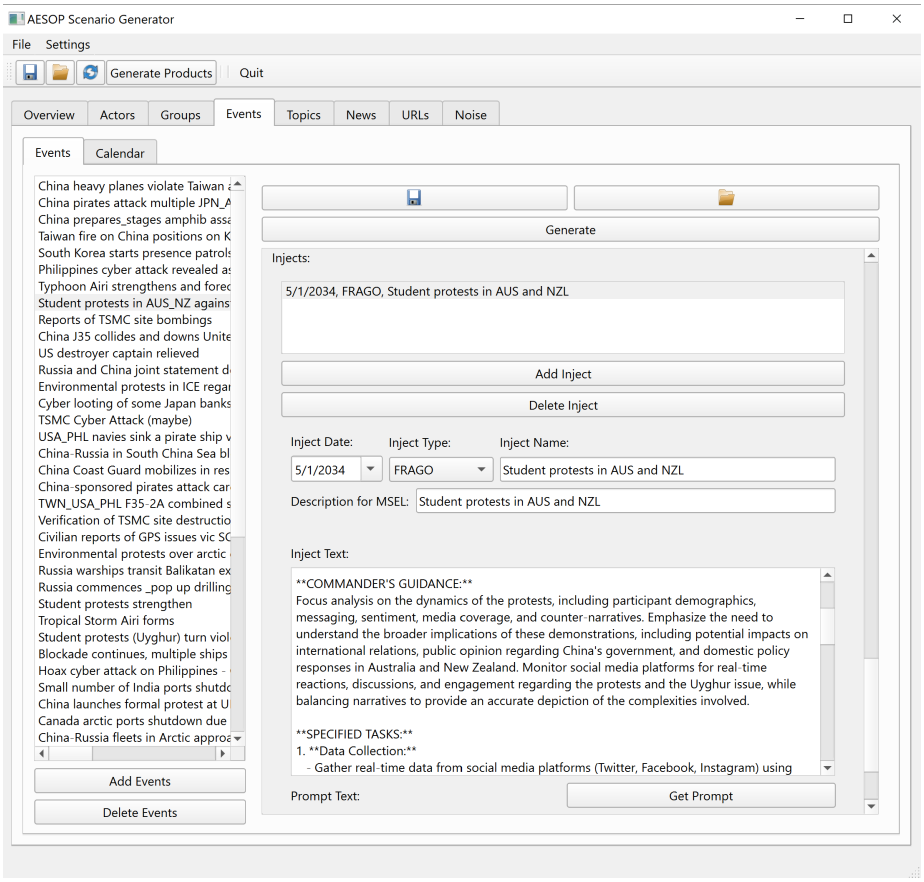

Figure 5.8: AESOP GUI for Inject creation.

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

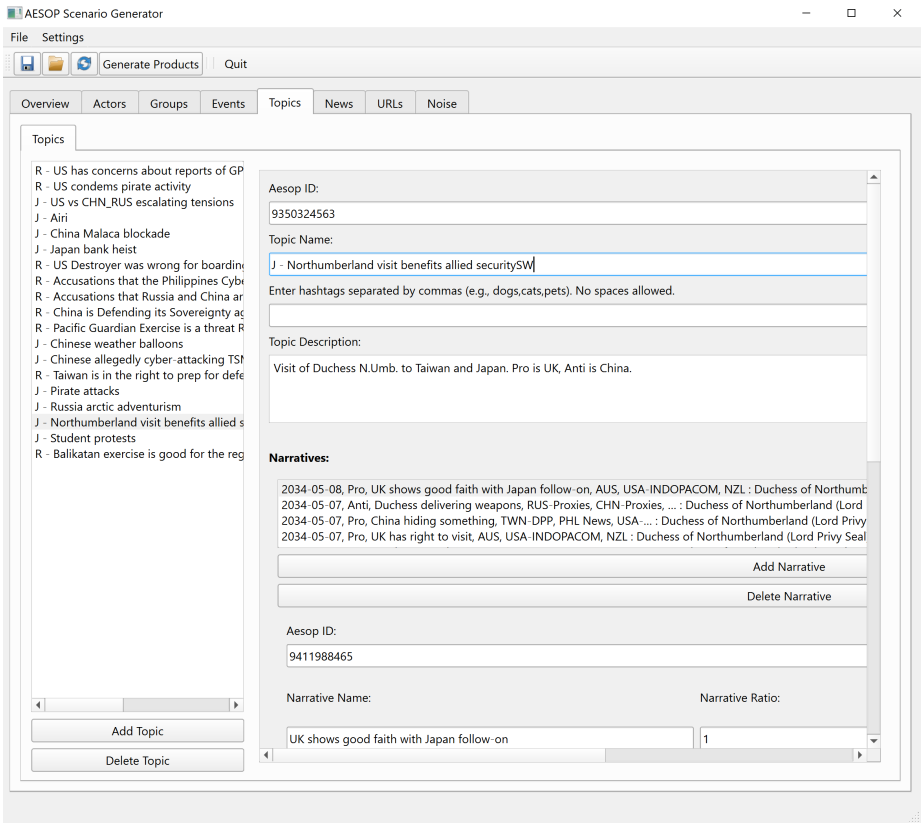

Figure 5.9: AESOP GUI for Topic creation.

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 propogate 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.

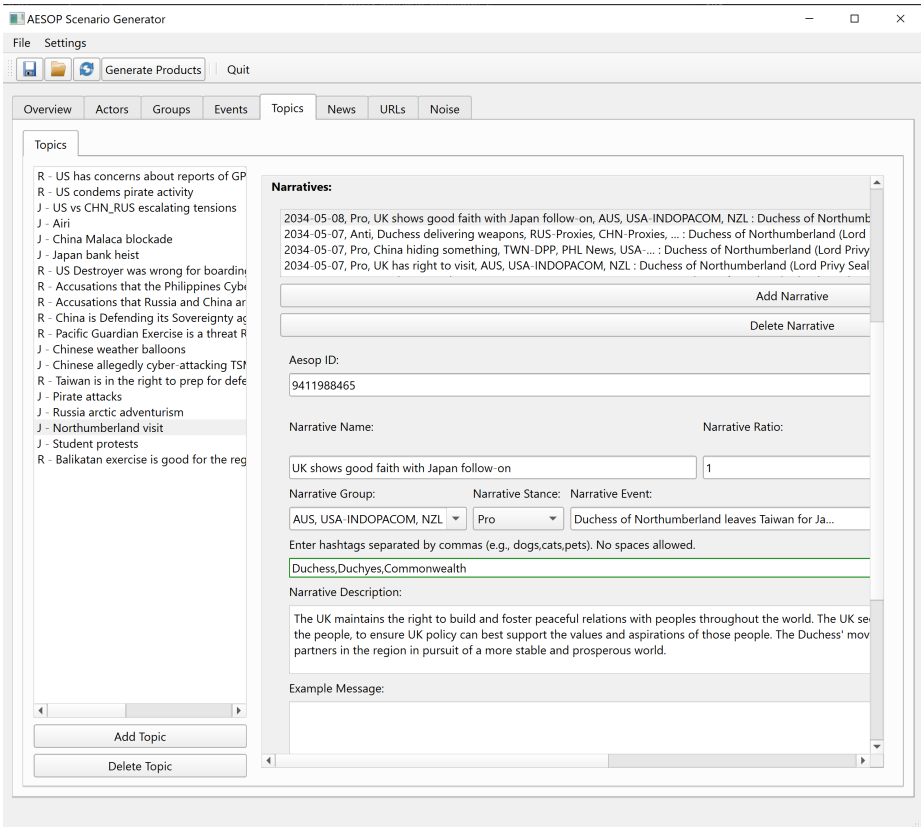

Figure 5.10: AESOP GUI for Narrative creation.

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

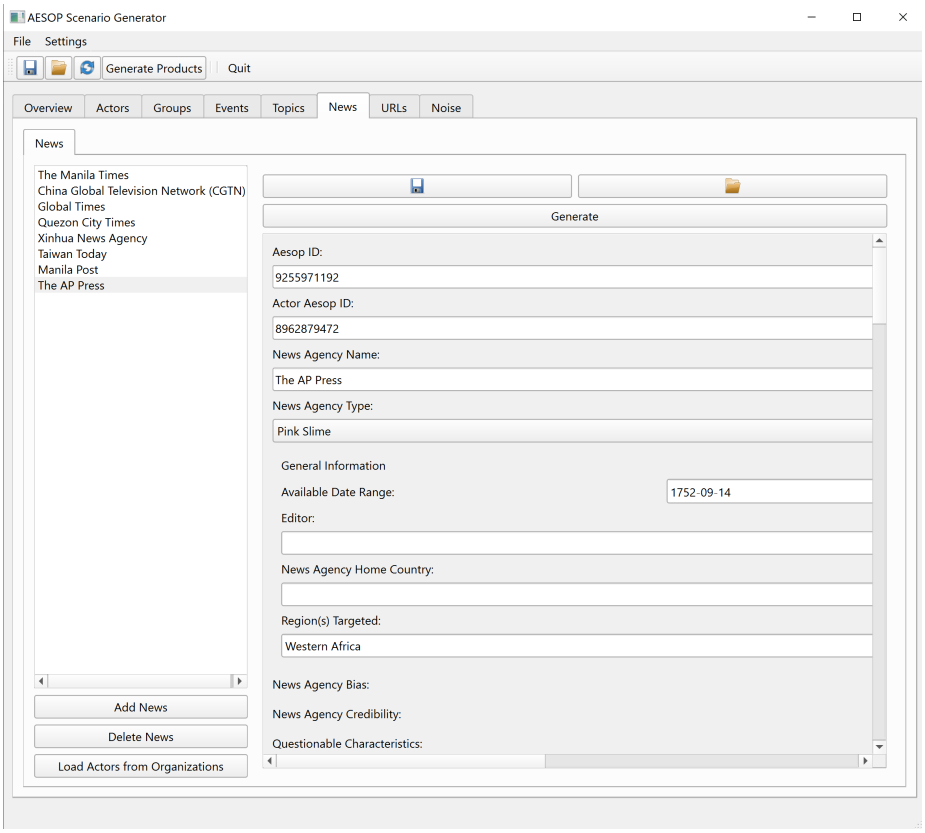

Figure 5.11: AESOP GUI for News Agency creation.

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.

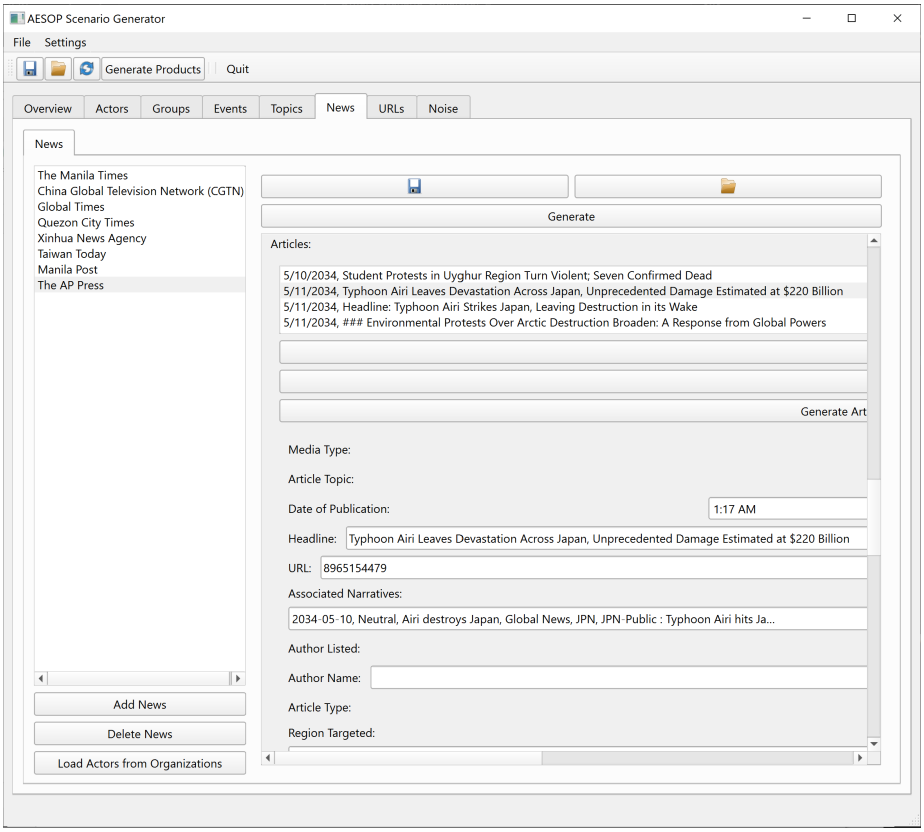

Figure 5.12: AESOP GUI for Article creation.

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

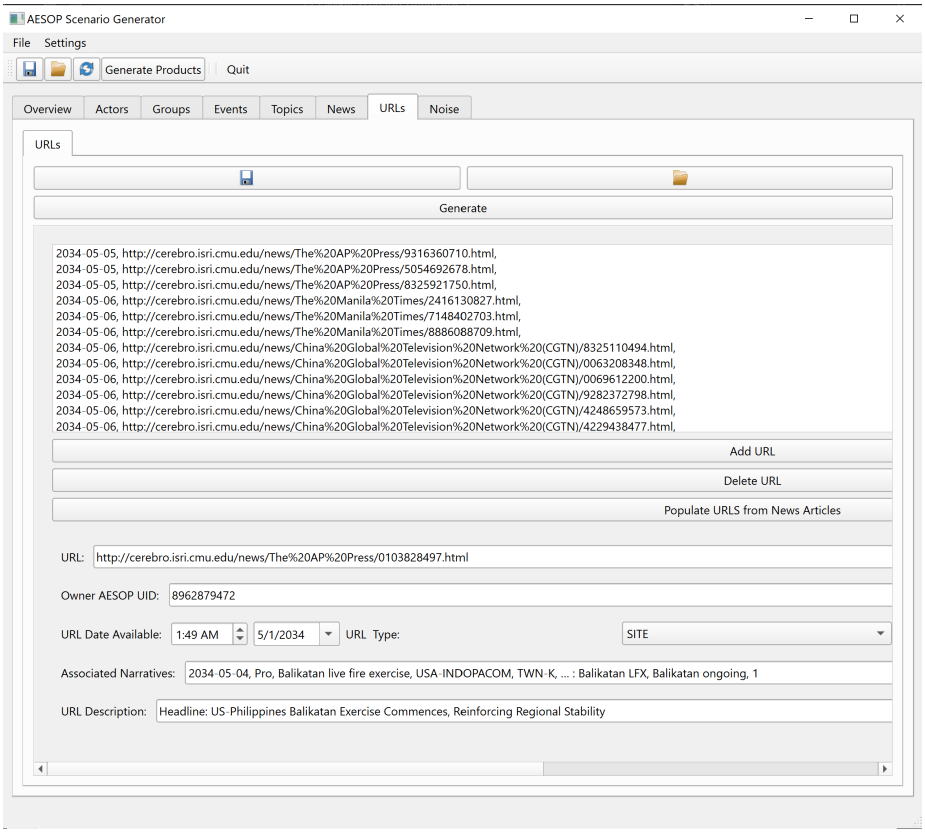

Figure 5.13: AESOP GUI for URL creation.

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

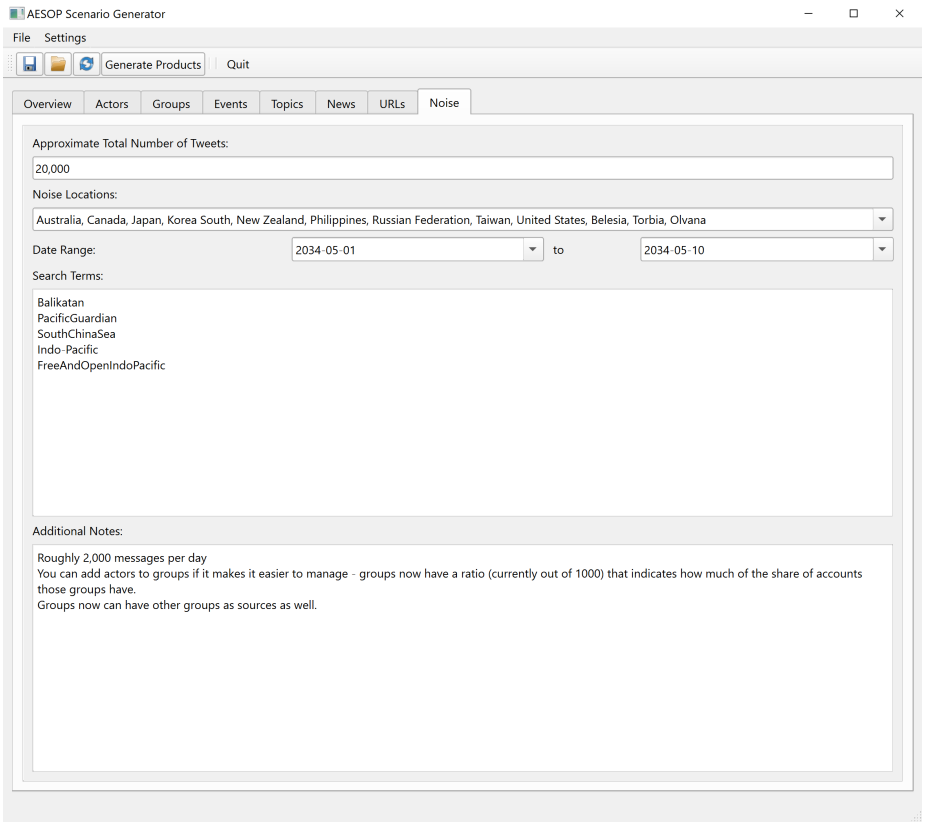

Figure 5.14: AESOP GUI for Noise creation.

5.4 AESOP Outputs

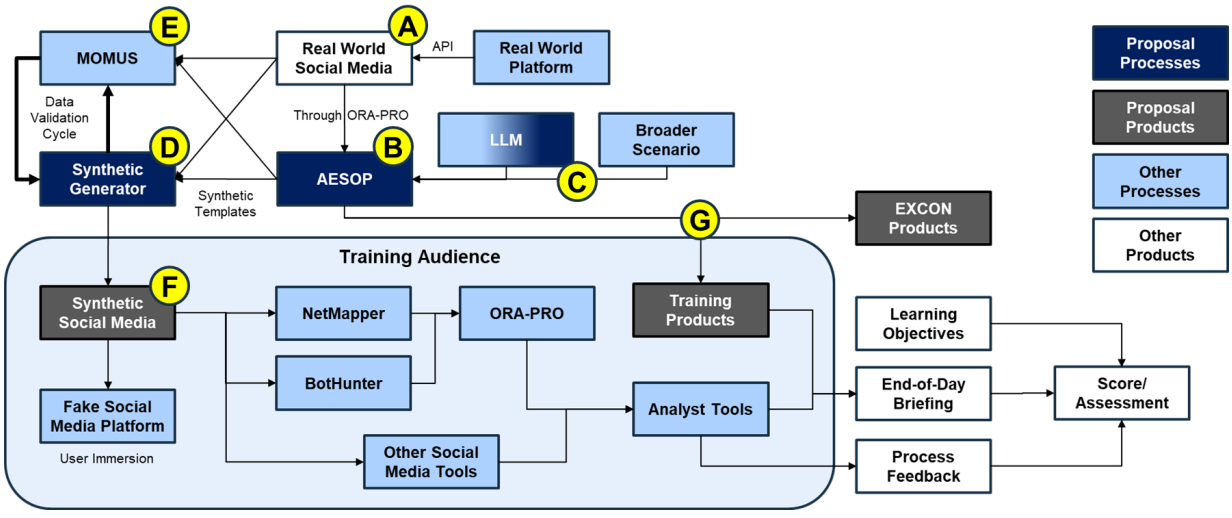

Figure 5.15: Project OMEN training flow.

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 B

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	https://oasis-open.github.io/cti-documentation/
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	https://www.w3.org/TR/activitystreams-core/
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	http://xmlns.com/foaf/spec/
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	https://web.archive.org/web/20220331224416/http://sioc-project.org/
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	US Intelligence Community (IC)	https://www.dni.gov/index.php/who-we-are/organizations/ic-cio/ic-technical-specifications/information-resource-metadata
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.[55] 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.[58]. 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.[64] 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.[20]. 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.[19] 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.[59] 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. [37] 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.

Table 5.2: AESOP Outputs and File Types

Feature	Archive File	Synthetic Template	EXCON	Participants
Agents				
People	JSON	JSON	DOCX	DOCX
Organizations	JSON	JSON	DOCX	DOCX
Bots	JSON	JSON	DOCX	DOCX
Telegram				
Account	JSON	JSON		
Channel	JSON	JSON		
Messages	JSON	JSON		
X/Twitter				
Account	JSON	JSON		
Messages	JSON	JSON		
Groups				
Group	JSON	JSON	DOCX	
Topics				
Topic	JSON	JSON	DOCX	
Narrative	JSON	JSON	DOCX	
Events				
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
News				
News Agency	JSON			
News Articles	JSON			
URLs				
URL	JSON	JSON		
Master Synch Event List				
Populated MSEL	JSON		XLSX	
Scenario Overview				
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.[54] [26] 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.[25] Interestingly, Change 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.[40] [62]

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,[25] the recreation of human social media text output,[67] or the generation of a specific set of synthetic training data.[68][39][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 [16] – is a model focused on Twitter interactions. The simulation accounts for both emotion and logic – ensuring

Table 6.1: Docking Lite

Feature	SynX	twitter_sim2.0	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	✓			

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[35]. 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.[25] 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.

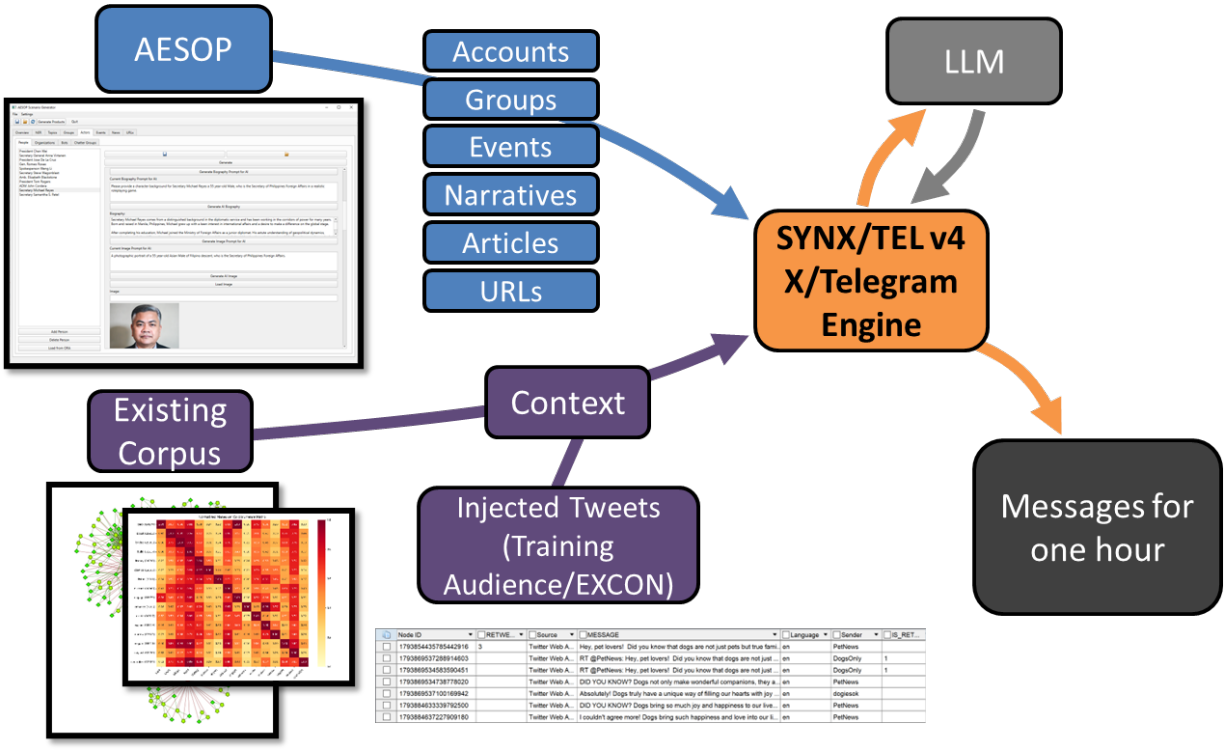

Figure 6.1: Overview diagram of SynTel/X

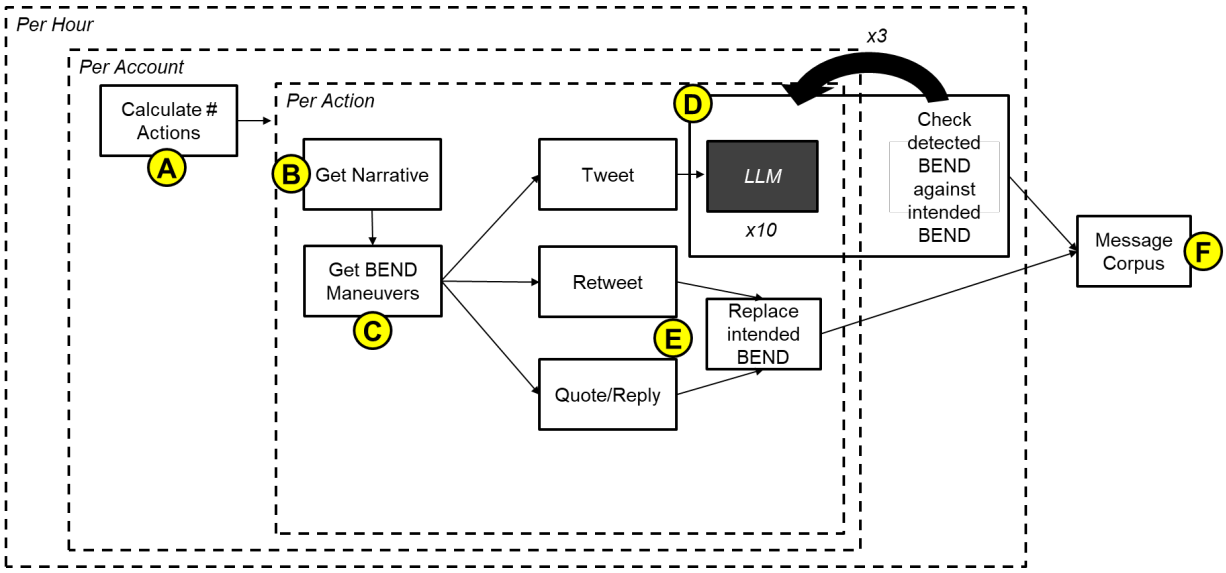

Figure 6.2: Example SynX logic flow.

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:

- $R_o = [r_{o,\text{min}}, r_{o,\text{max}}]$ be the range of original tweets per day
- $R_r = [r_{r,\text{min}}, r_{r,\text{max}}]$ be the range of retweets per day
- $R_q = [r_{q,\text{min}}, r_{q,\text{max}}]$ 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:

$$\begin{aligned}
\text{min_posts} &= r_{o,\text{min}} + r_{r,\text{min}} + r_{q,\text{min}} \\
\text{max_posts} &= r_{o,\text{max}} + r_{r,\text{max}} + r_{q,\text{max}} \\
\text{adjusted_min} &= \lfloor \text{min_posts} \times E \rfloor \\
\text{adjusted_max} &= \lfloor \text{max_posts} \times E \rfloor
\end{aligned}$$

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 $h \in \{0, 1, \ldots, 23\}$ 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:

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

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:

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

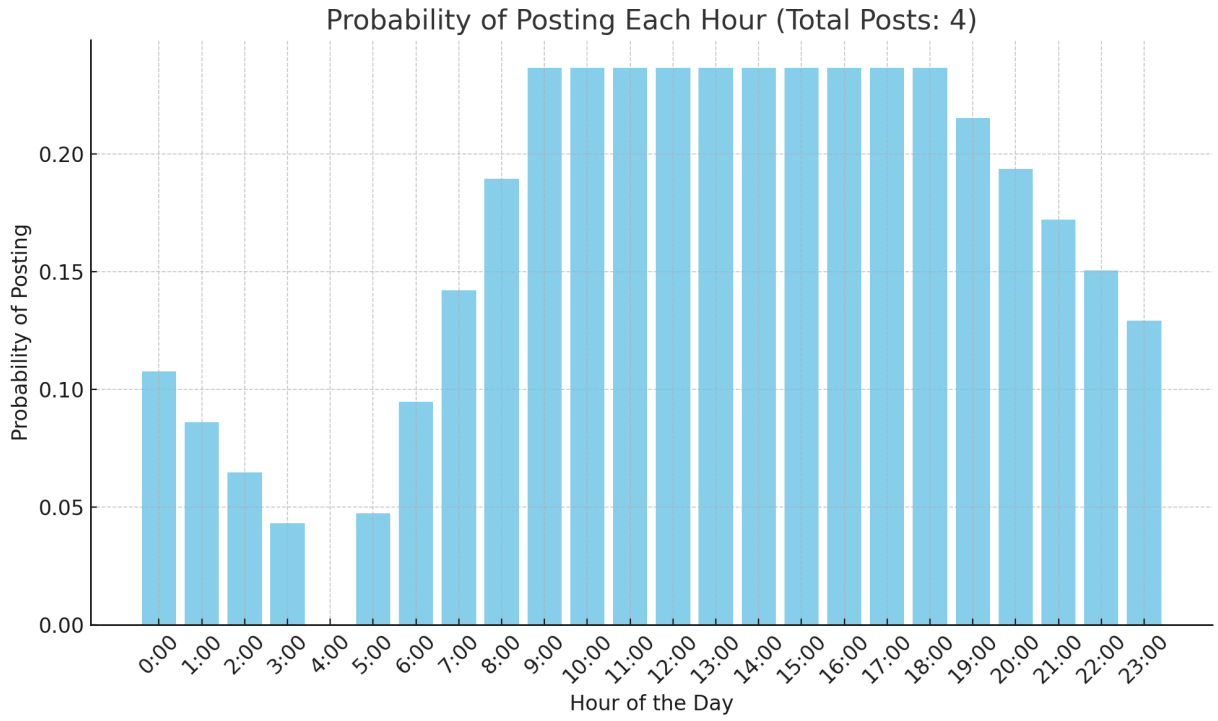

Figure 6.3: Graph of probability of posting each hour for a user active 0900 to 1800 who posts 4 times a day.

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
- $\mathcal{G}(A) = \{G_1, G_2, \ldots, G_m\}$ 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:

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

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

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

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:

- $M = \{M_1, M_2, \ldots, M_{16}\}$ 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:

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

From this seed BEND maneuver we can use a Markov chain to determine the total BEND maneuvers for the action. Or, more precisely, we can use a generalization of a Markov-like model that includes memory and depends upon the full history set of maneuvers.

BEND Sequence Generation from a Start Maneuver

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

1. Initial Maneuver

$$S_0 = [m_0]$$

2. At Each Step

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

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

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

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

3. Mask Already Chosen Maneuvers

To ensure uniqueness, set:

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

4. Normalize Probabilities

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

5. Sampling

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

6. Termination Condition

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

From this we arrive at our full sequence of BEND maneuvers: $S = [m_0, m_1, \ldots, m_k]$. 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.[25] 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

	Derivative	**Mentions**	**URLs**
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 F

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 asks 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:

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

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

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

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.

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

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:

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

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 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

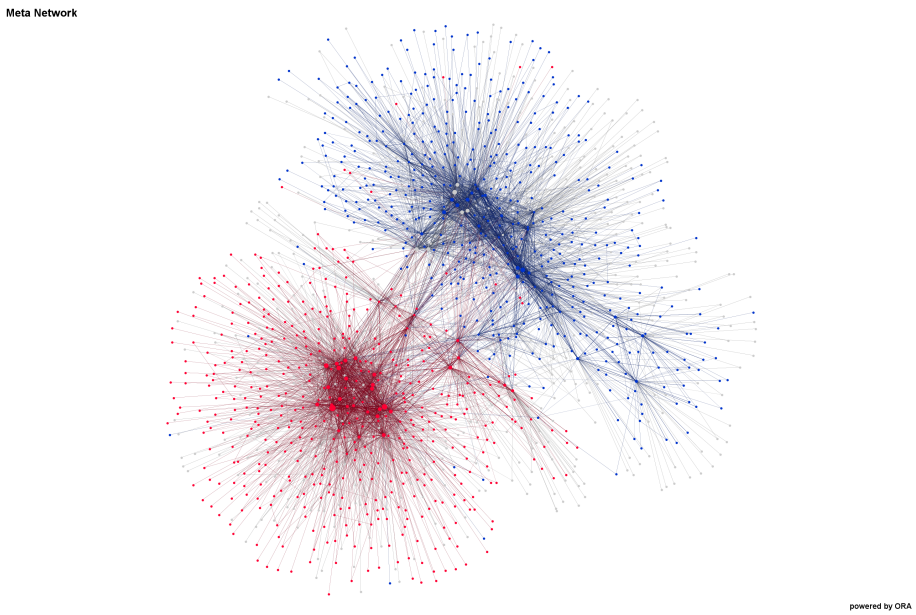

Figure 6.4: An output network from SynX with pro-US stance detection run through ORA-Pro. Blue nodes are pro-US actors and red nodes are anti-US actors with gray nodes being neutral. Nodes are sized by total degree centrality.

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 ingroup continuously. Both of these types of bots operate similarly to those seen by Ng and Carley in 2023[56]. 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

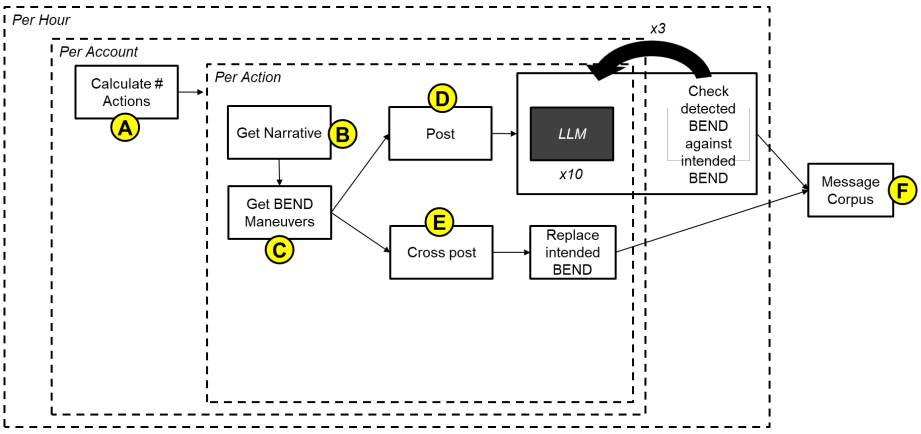

Figure 6.5: Telegram generator message creation flow. Dark gray parallelograms are synthetic templates provided by AESOP and referenced by the generator.

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

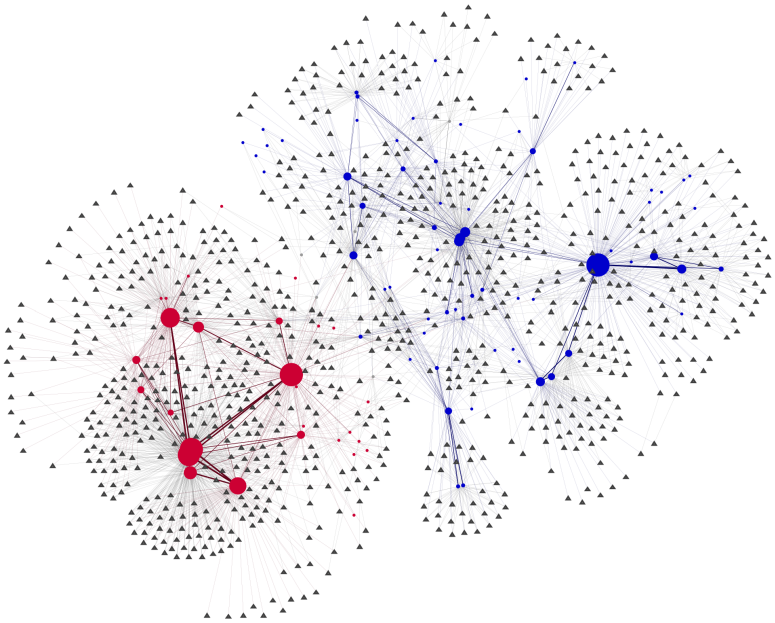

Figure 6.6: An output network from SynTel with pro-US stance detection run through ORA-Pro. Blue nodes are pro-US actors and channels. Red nodes are anti-US actors and channels Triangle shaped nodes are messages. Nodes are sized by total degree centrality.

suggested by Chang et al. in 2024 during their development of an LLM-based synthetic network simulation.[25] 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 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
Attentions spans limit how many users are affected	Reach is determined by leadership	Kang and Lerman, 2013. [49] Lu et al., 2014. [53]
Commonality of reposts	High level of derivative messages (50+%)	Beskow and Carley's "Agent Based Simulation of Bot Disinformation Maneuvers in Twitter" from 2019.[12]
Which messages are being reposted	Power law distribution	Lu et al., 2014. [53]
Real world likelihood of BEND maneuvers	BEND maneuver Markov-like model from 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

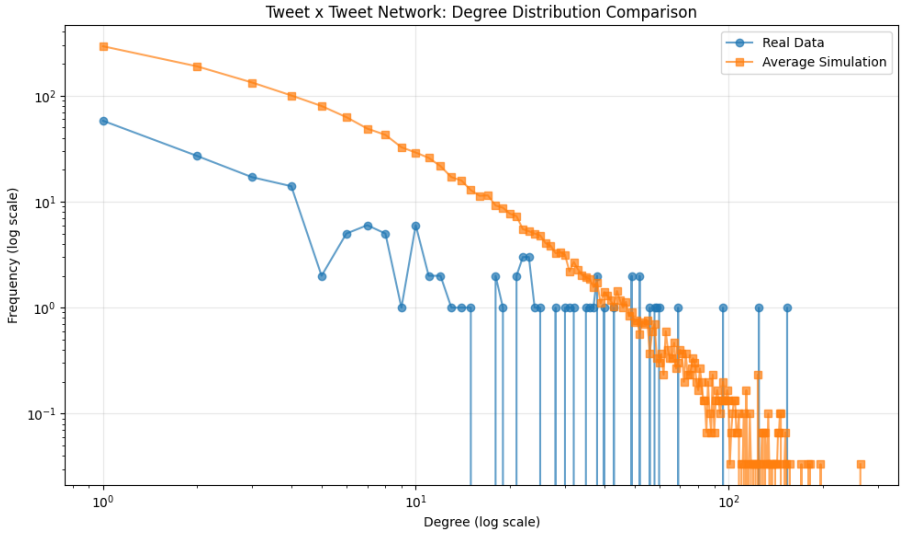

Figure 6.7: Both the real data and the SynX display appropriate power-law properties for the Tweet x Tweet (Retweet) network

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 doin 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

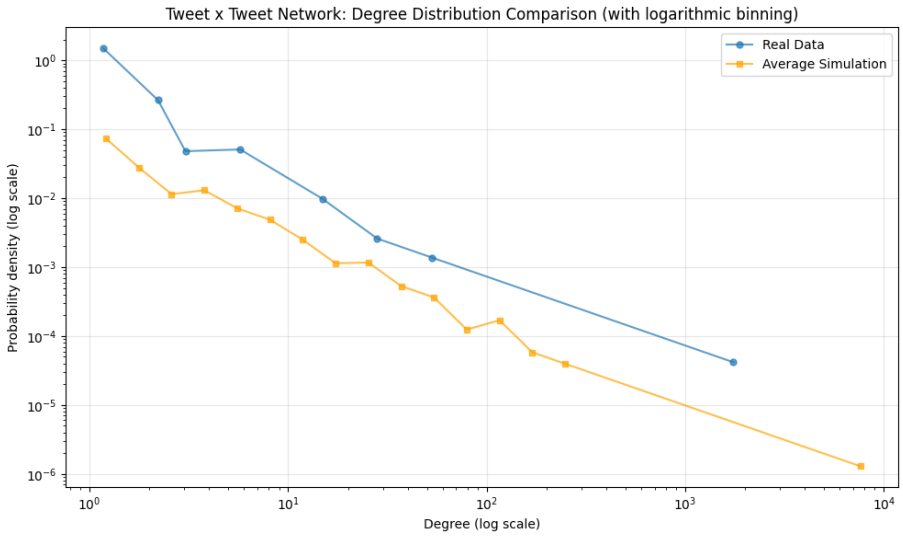

Figure 6.8: Binning is not always a good idea, but with the discontinuous data from the real data this chart helps depict the form convergence between the real and synthetic data sets

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	8 cells	
Replications per cell	30	
180 total runs (108,000+ messages)		

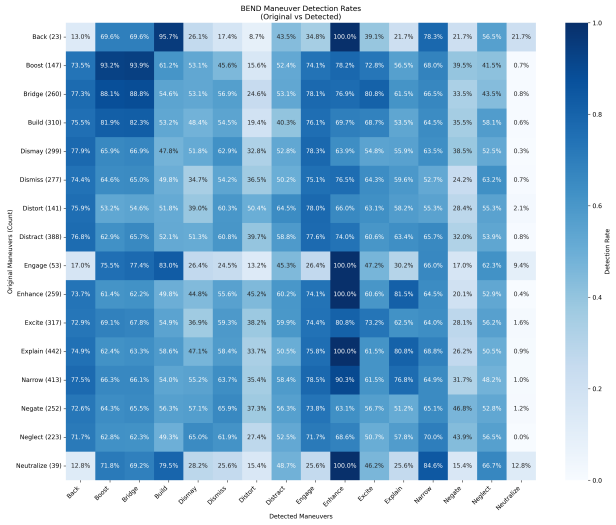

Figure 6.9: Example co-occurrence heatmap (that shows intended BEND maneuvers mapped to detected BEND maneuvers.

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	**0.35**$^{+}$	**-0.42**$^{-}$	**-0.26**$^{-}$	**-0.22**$^{-}$
boost	**-0.67**$^{-}$	**-0.89**$^{-}$	**-0.69**$^{-}$	**-0.66**$^{-}$
bridge	**-0.44**$^{-}$	**-0.38**$^{-}$	-0.10	-0.05
build	-0.24	-0.07	0.17	0.21
dismay	**-0.87**$^{-}$	**-0.80**$^{-}$	**-0.35**$^{-}$	**-0.50**$^{-}$
dismiss	-0.21	-0.11	**0.17**$^{+}$	0.11
distort	**-0.71**$^{-}$	**-0.65**$^{-}$	-0.29	**-0.31**$^{-}$
distract	**-0.73**$^{-}$	**-0.64**$^{-}$	**-0.31**$^{-}$	**-0.34**$^{-}$
engage	-0.12	**-0.47**$^{-}$	**-0.28**$^{-}$	**-0.24**$^{-}$
enhance	0.11	0.17	**0.45**$^{+}$	**0.39**$^{+}$
excite	**-0.64**$^{-}$	**-0.71**$^{-}$	**-0.39**$^{-}$	**-0.44**$^{-}$
explain	**-0.46**$^{-}$	**-0.46**$^{-}$	0.04	-0.06
narrow	**-0.41**$^{-}$	**-0.30**$^{-}$	0.13	0.13
negate	**-1.00**$^{-}$	**-1.08**$^{-}$	**-0.70**$^{-}$	**-0.76**$^{-}$
neglect	**0.98**$^{+}$	**0.95**$^{+}$	**1.22**$^{+}$	**1.18**$^{+}$
neutralize	**-0.33**$^{-}$	**-0.63**$^{-}$	**-0.50**$^{-}$	**-0.44**$^{-}$

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.

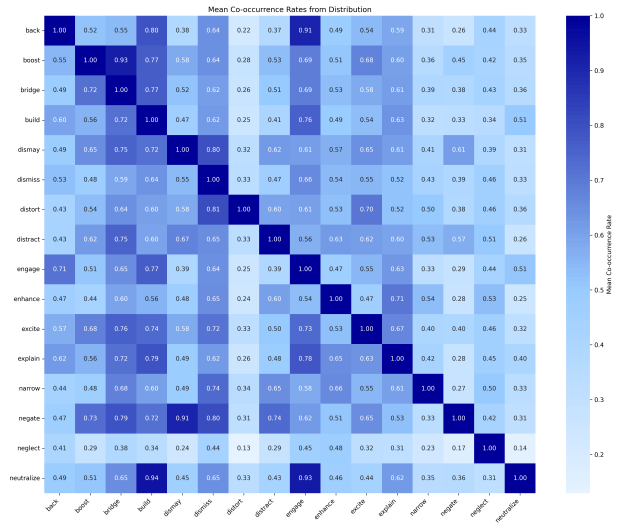

Figure 6.10: Co-occurrence heatmap of the Balikatan 22 data - note that there is no concept of intended vs detected BEND maneuvers in real data

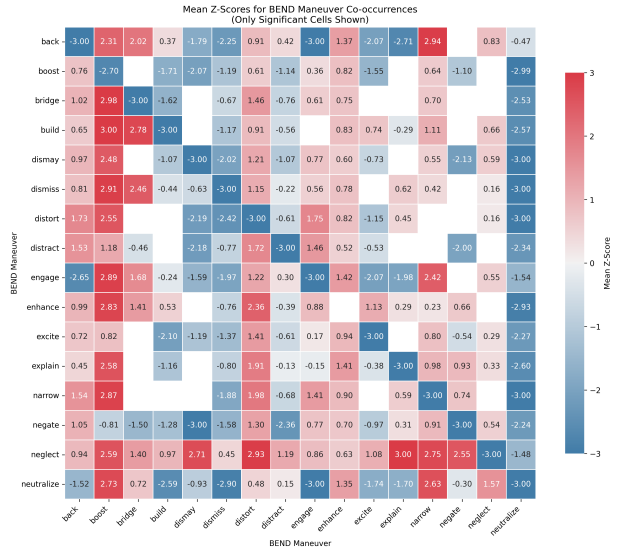

Figure 6.11: Co-occurrence matrix across all runs that had the BEND definitions in the system prompt but did not repoll the LLM

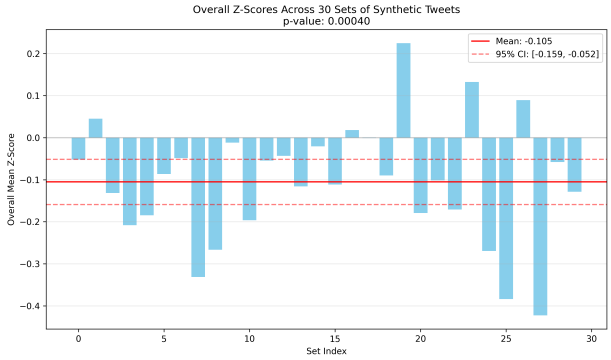

Figure 6.12: A comparison of the overall distance from the mean for all runs that had the BEND definitions in the system prompt but did not repoll the LLM

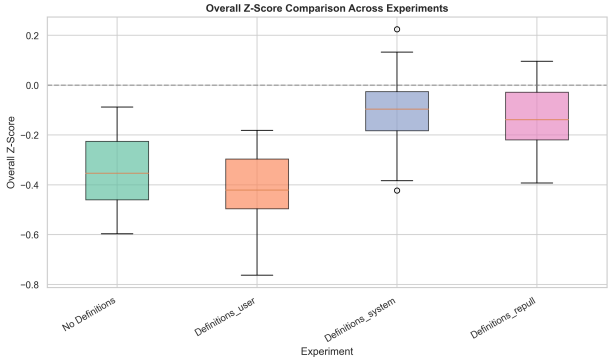

Figure 6.13: A comparison of the overall distance from the mean across all BEND maneuvers from each of the four major experiment setups

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

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

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 theses 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.[14] 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 upon 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 strategic bombing in World War II, rather than looking at a B-17 and its payload and determining that this will be a firebombing mission on Dresden, I want to look at the burned out city to assess not only the action taken but also the effectiveness of that action. I want to be able to look at Schweinfurt, see that the ball-bearing factory is destroyed, and point to a bombing group action that occurred on a particular night. I could not tell you which bomb destroyed the factory or even which bomber, but I can definitely point to a specific raid. In the same way, I will not be able to point to a specific message or actor, but I will be able to identify a narrative campaign associated with an effect induced in the target.

This will require bounded, over-time comparisons of groups in order to detect changes in the 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?

I have developed a set of network and narrative metrics to define BEND 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. The definitions below are derived from Blane[14] and paired with their corresponding effects-based metric.

TOG: Topic-Oriented Group, identified by ANDing the Agent x Agent (strong ties) network and the Agent x Agent (concept) network and performing Leiden TOG community: the linked Topic Oriented Group of a cluster in time 1 with a cluster in time 2 TOG cluster: an identified Topic Oriented Group in a single time (two TOG clusters make up a TOG community)

Back maneuvers have discussion or actions that increase the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic. 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.

Negate maneuvers include discussion or actions that decrease the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic. Therefore, 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.

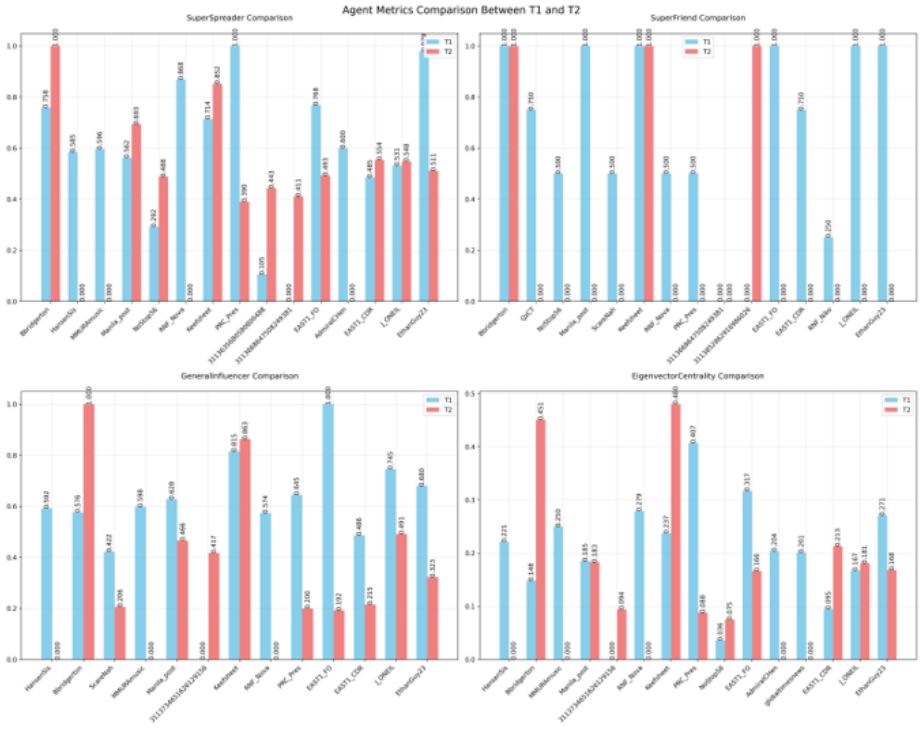

Figure 7.1: Actual results of back/negate.

Build maneuvers are marked by discussion or actions that create a group, or the appearance of a group, where there was none before. Therefore, new group emergence is required for this maneuver - agent interactions within the group should change positively over time more than in the baseline corpus.

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. Effects-based metrics will be the appearance of multiple groups where only one was present before within the actor network or the disappearance of links on the bipartite network from meta-agent group node to topic/stance nodes.

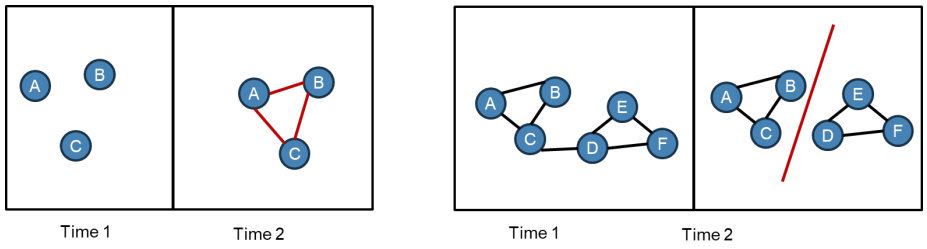

Figure 7.2: Visualization of back/negate.

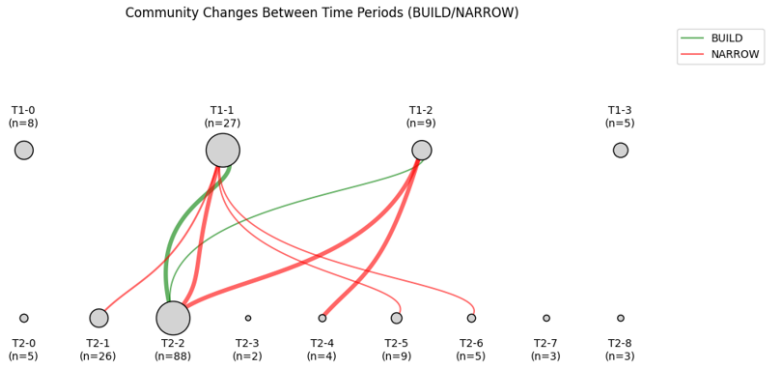

Figure 7.3: Actual results of back/negate.

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 both centrality and betweenness of the edge nodes of two groups. A positive change over time above the baseline corpus indicates a bridge maneuver.

Neutralize maneuvers have discussion or actions that cause a group to be, or appear to be, no longer of relevance, or the group is dismantled. I will be looking for target group nodes that have more in common (connections) with other groups than themselves (group disappears) over the time-frame observed.

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 group size and graph density for positive changes over time above the baseline corpus.

Neglect maneuvers show discussion or actions that decrease the size of a group and/or the connections among group members, or the appearance of such. For effects-based detection of

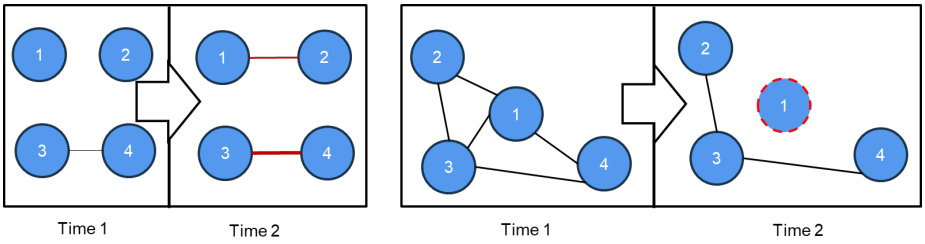

Figure 7.4: Visualization of bridge/neutralize.

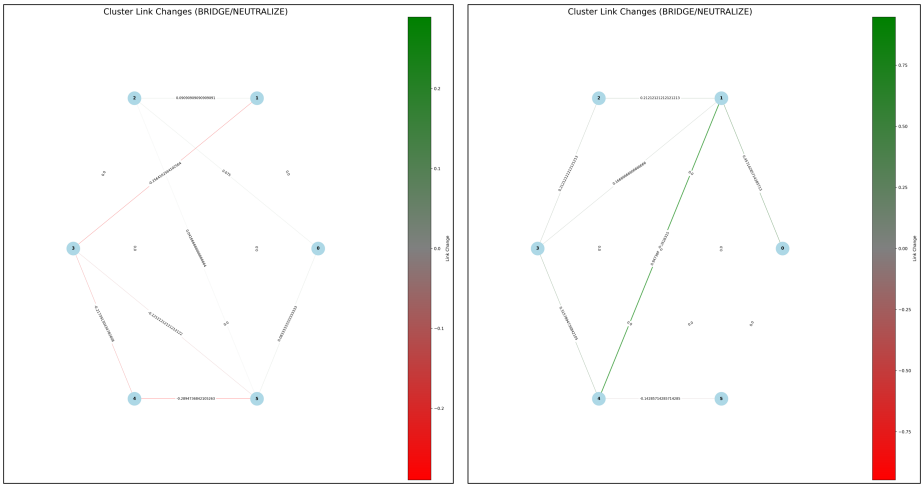

Figure 7.5: Actual results of bridge/neutralize.

the maneuver, I will look for density and/or size of a target group changing negatively over time above the baseline corpus.

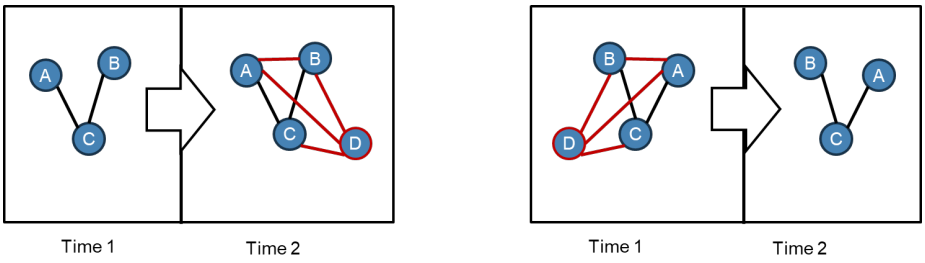

Figure 7.6: Visualization of boost/neglect.

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 and surprise over time above that of the baseline corpus.

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

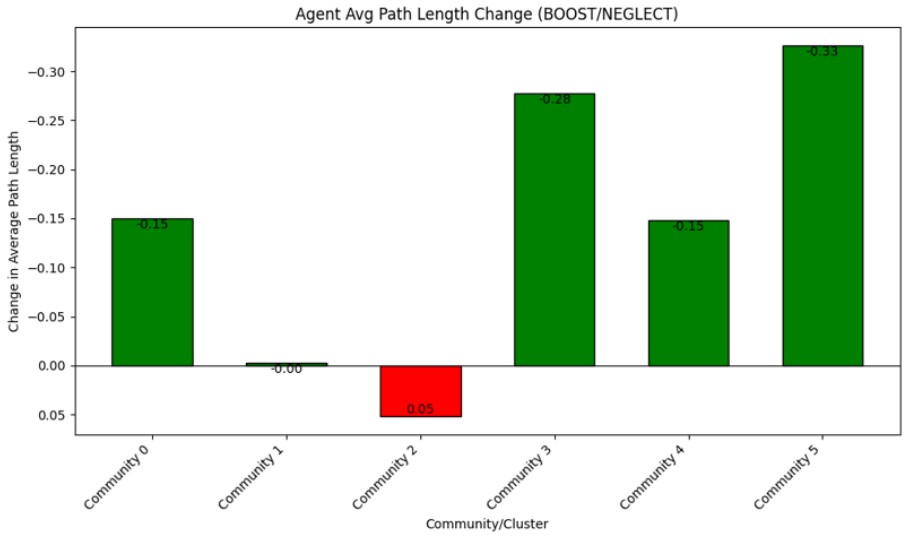

Figure 7.7: Actual results of boost/neglect.

anger, sadness, fear over time more than the baseline corpus.

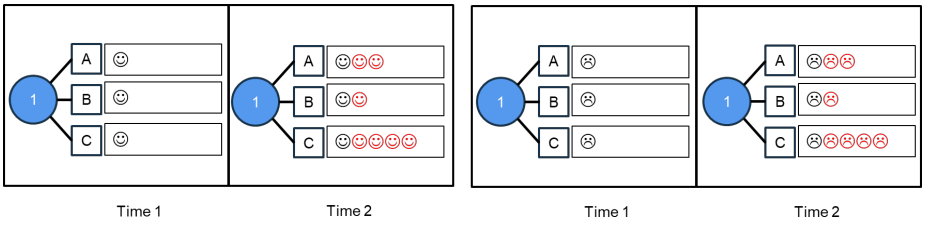

Figure 7.8: Visualization of excite/dismay.

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, I will look at topic specialization – with additional jargon and a net shift towards similar stance over time above the baseline being indicative of an explain maneuver.

Distort maneuvers include 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. This should induce in the target increased topic specialization – additional jargon in message and net shift towards opposite stance over time more than the baseline corpus.

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 the topic with the target group above the baseline corpus.

Dismiss maneuvers are marked by 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. I will be looking for a negative change over time in the topics

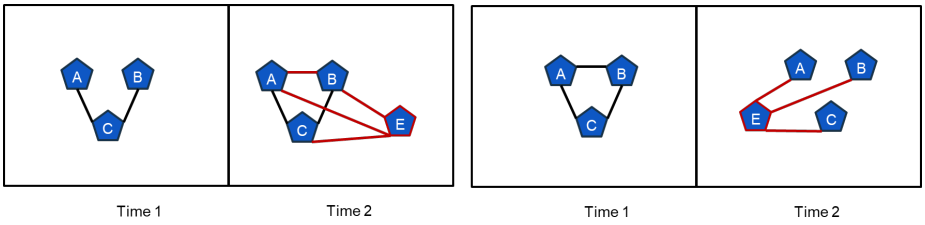

Figure 7.9: Visualization of explain/distort.

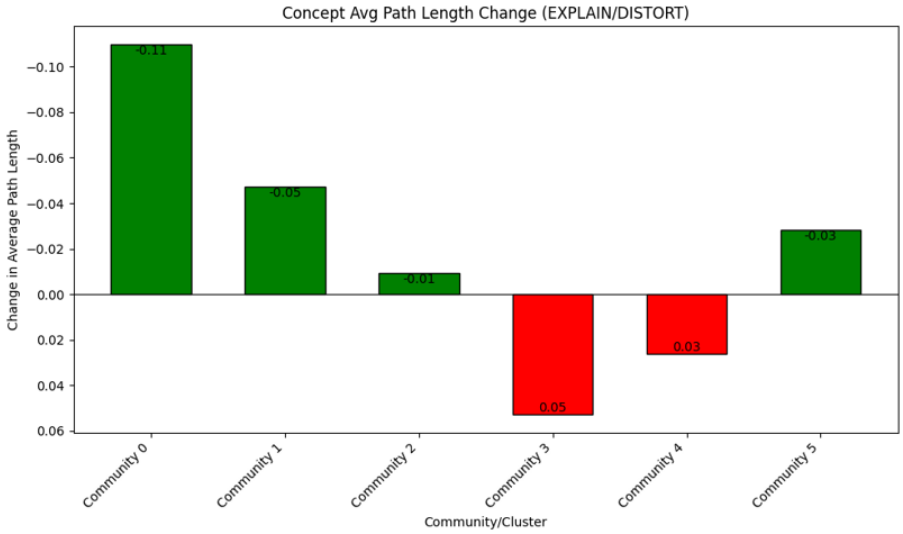

Figure 7.10: Actual results of explain/distort.

proportional representation - greater in magnitude than the in baseline corpus.

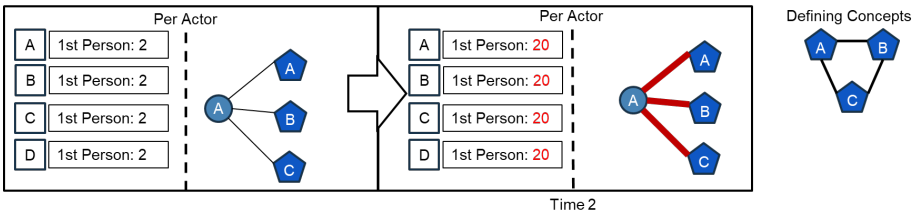

Figure 7.11: Visualization of engage/dismiss.

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. Effects-based metrics will be increased linkages (density) and centrality or betweenness changing positively over time above the baseline corpus.

Distract maneuvers require 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. For this, I will look for decreased linkage/density and decreased centrality and betweenness over time - in greater magnitudes than experienced by the baseline corpus.

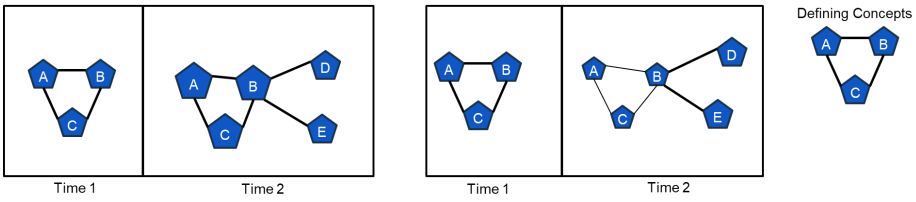

Figure 7.12: Visualization of enhance/distract.

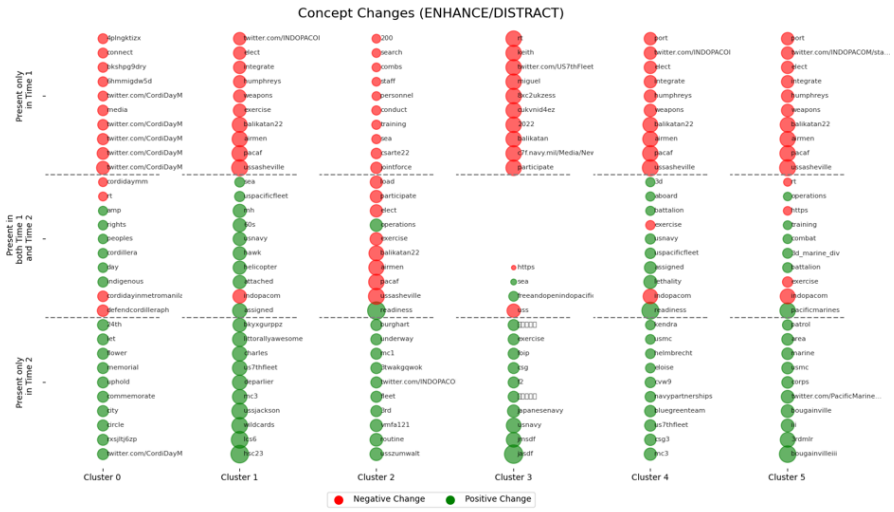

Figure 7.13: Actual results of enhance/distract.

A summary of the maneuvers - their definitions and effects-based metrics can be found in Fig. 7.14.

Studies

7.2 Methods

7.3 Results

7.4 Implications

7.5 Conclusions

Name	Definition	Effects-Based Detection
Back	Discussion or actions that increase the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic	Centrality in interaction network, importance to group/ changes positively over the time more than the baseline corpus
Build	Discussion or actions that create a group, or the appearance of a group, where there was none before	New group – agent interactions / changes positively over the time more than the baseline corpus
Bridge	Discussion or actions that build a connection between two or more groups or create the appearance of such a connection	Centrality, betweenness of the edge nodes of two groups / changes positively over the time more than the baseline corpus
Boost	Discussion or actions that increase the size of a group and/or the connections among group members, or the appearance of such	Size of group, graph density / changes positively over the time more than the baseline corpus
Excite	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	Target packet emotional valence will be higher in happiness and surprise / changes positively over the time more than the baseline corpus
Explain	Discussion or actions that clarify a topic to the targeted community or actor often by providing details on, or elaborations on, the topic	Topic specialization– additional jargon and net shift towards same stance / changes over the time more than the baseline corpus
Engage	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	Topics proportional representation / changes positively over the time more than the baseline corpus
Enhance	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	Increased linkage and centrality, betweenness / changes positively over the time more than the baseline corpus
Negate	Discussion or actions that decrease the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic	Centrality of node / changes negatively over the time more than the baseline corpus
Neutralize	Discussion or actions that cause a group to be, or appear to be, no longer of relevance, e.g., because it was dismantled	Group nodes have more in common with other groups than themselves (group disappears)
Narrow	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	Multiple groups where only one was present before, fewer links on bipartite network from meta-agent group node to topic/stance nodes
Neglect	Discussion or actions that decrease the size of a group and/or the connections among group members, or the appearance of such	Density and/or size / changes negatively over the time above the baseline corpus
Dismay	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	Target packet emotional valence will be higher in anger, sadness, fear / changes over the time more than the baseline corpus
Distort	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	Topic specialization– additional jargon and net shift towards opposite stance / changes over the time more than the baseline corpus
Dismiss	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	Topics proportional representation / changes negatively over the time more than the baseline corpus
Distract	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	Decreased linkage and centrality, betweenness / changes over the time more than the baseline corpus

Figure 7.14: BEND definitions to effects mapping.

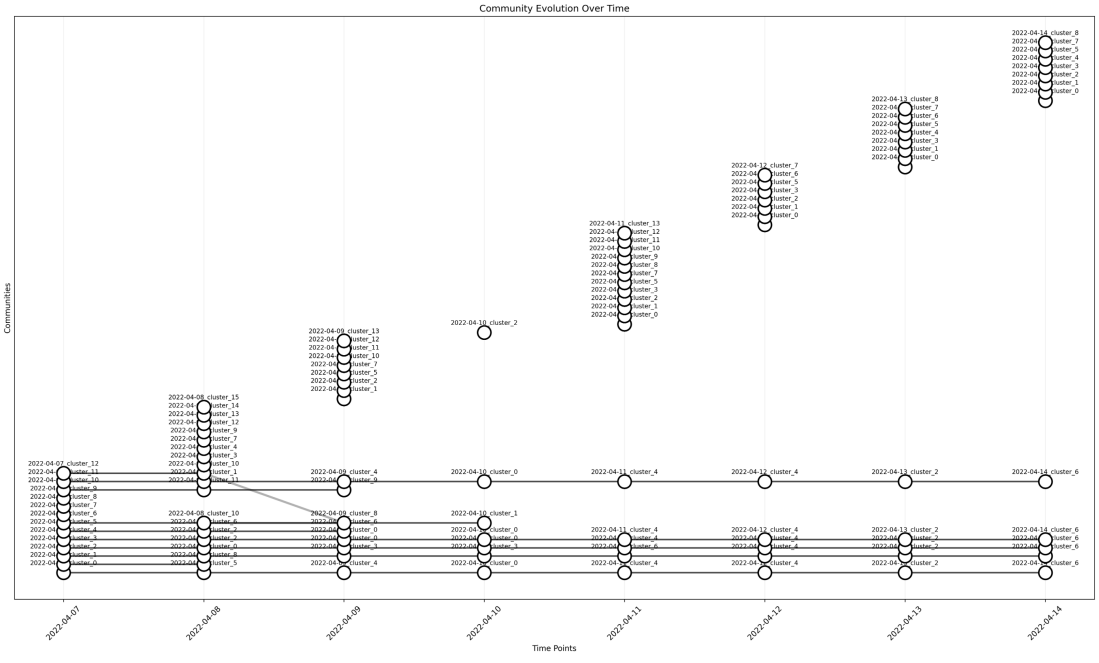

Figure 7.15: Balikatan Groups.

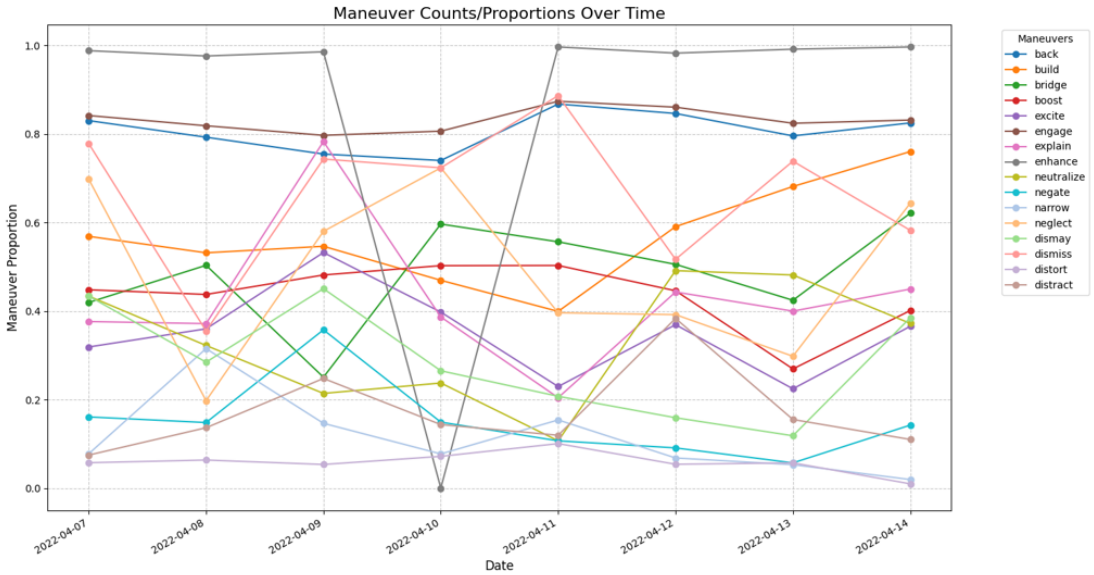

Figure 7.16: Balikatan BEND Counts.

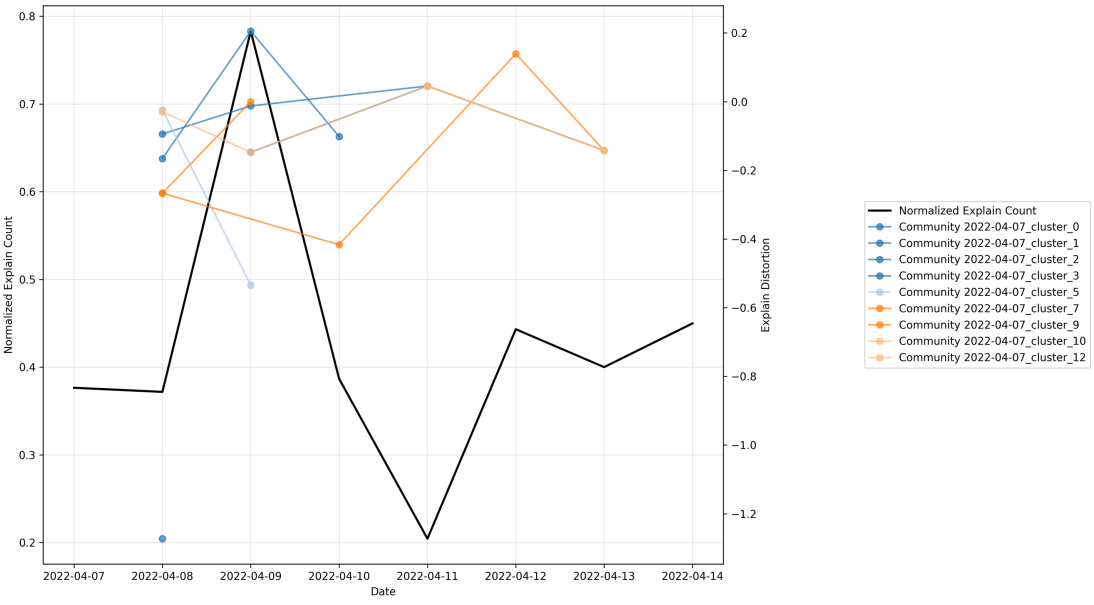

Figure 7.17: Balikatan effect vs maneuver.

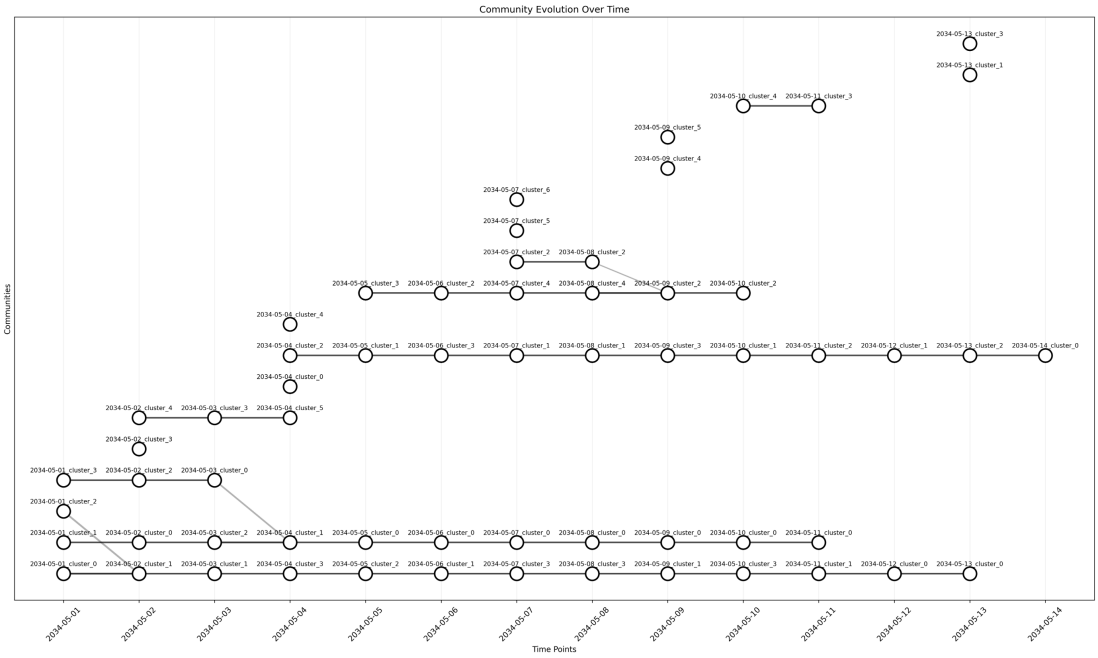

Figure 7.18: ROF Groups.

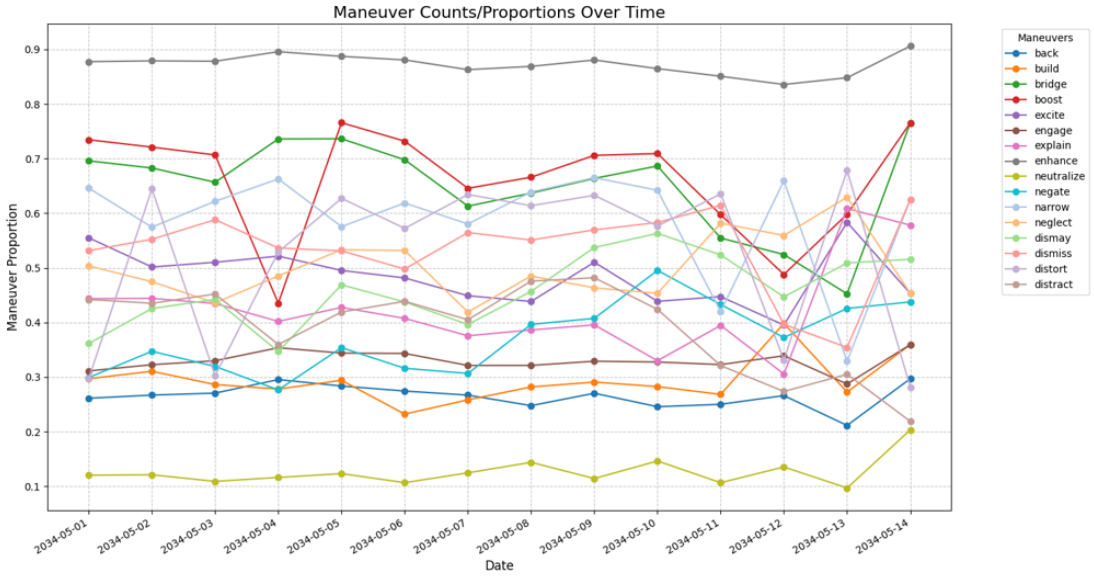

Figure 7.19: ROF BEND Counts.

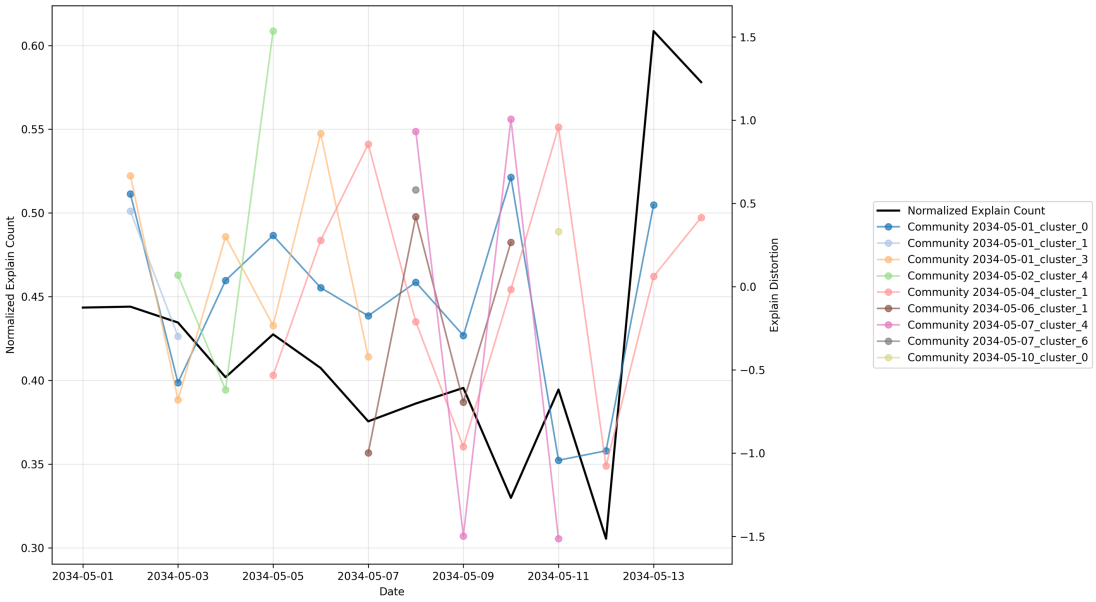

Figure 7.20: ROF effect vs maneuver.

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 1385*. 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://www.doctrine.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://www.doctrine.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] The official python library for the OpenAI API. URL https://github.com/openai/openai-python. original-date: 2020-10-25T23:23:54Z. 2.2.1, 5.3

[8] 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. 4.2

[9] 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

[10] 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

[11] 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

[12] 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

[13] David M. Beskow and Kathleen M. Carley. Social cybersecurity: an emerging national security requirement. 99(2):117–127, . URL https://apps.dtic.mil/sti/citations/AD1108494. 3.1

[14] 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. 2.1, 2.1, 3.1, 3.1, 7.1

[15] 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

[16] Janice T. Blane, J. D. Moffitt, and Kathleen M. Carley. Simulating social-cyber maneuvers to deter disinformation campaigns. In Robert Thomson, Muhammad Nihal Hussain, 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. 6.2

[17] 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

[18] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. 3:993–1022. 5.3

[19] 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

[20] Dan Brickley and Libby Miller. Foaf vocabulary specification 0.99. http://xmlns.com/foaf/spec/, 2014. Accessed: 2025-03-13. 5.4

[21] 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

[22] 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 http://link.springer.com/10.1007/978-3-319-93372-6_42. 3.1

[23] 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

[24] *Joint Information Operations Proponent*. Chairmain of the Joint Chiefs of Staff, Washington, DC, USA, 2014. URL https://www.jcs.mil/Portals/36/Documents/Library/Instructions/3210_01.pdf. 4.2

[25] 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

[26] 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

[27] 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

[28] 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

[29] *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

[30] *Information in Air Force Operations*. Department of the Air Force, Washington, DC, USA, 2018. URL https://www.doctrine.af.mil/Portals/61/documents/AFDP_3-13/3-13-AFDP-INFO-OPS.pdf. 3.1, 4.2

[31] *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. 4.2

[32] *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

[33] *Army Information*. Department of the Army, Washington, DC, USA, 2023. URL https://irp.fas.org/doddir/army/adp3-13.pdf. 4.2

[34] *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. 4.2

[35] Stephen Dipple and Kathleen M. Carley. *Construct User Guide*. URL https://www.cmu.edu/casos-center/publications/cmu-s3d-23-104.pdf. 6.2

[36] Geoffrey B. Dobson and Kathleen M. Carley. Cyber-FIT: An agent-based modelling ap-

proach 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. 3.1

[37] 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

[38] 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. 3.1

[39] 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. 5.1, 6.2

[40] 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

[41] *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

[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. 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. 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. 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. 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-a 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. 6.5, 6.3, 6.5.1

[53] 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

[54] 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. 6.2

[55] Randall Munroe. Standards. https://xkcd.com/927/, 2011. Accessed: 2025-03-13. 5.4

[56] 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

[57] 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-t
3.1

[58] OASIS Open. Cti documentation. https://oasis-open.github.io/cti-documentation/. Accessed: 2025-03-13. 5.4

[59] 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

[60] 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

[61] 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. 3.1

[62] 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. *arXiv preprint arXiv:2304.03442*, 2023. URL https://arxiv.org/pdf/2304.03442. 6.2

[63] 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

[64] 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

[65] 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

[66] 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

[67] 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. 6.2

[68] 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

[69] 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

US Military BEND Products

Appendix B

AESOP Products

Appendix C

Data Standard

Appendix D

SynX Paper

Appendix E

BEND Effects Paper

Appendix F

SynX Prompts