AMITT Design Guide

Contents

1 Introduction

What this report covers, and who's responsible for that.

1.1 About this Report

This report covers the motivations, philosophy, use and intentions for the AMITT disinformation models. It's one of a series of reports on how information security principles and practices can be used to improve our understanding of cognitive security and improve responses to information operations, and specifically to disinformation campaigns and incidents.

The original brief for the report was to create a 1-2 page description of each tactic and technique in AMITT (the 'squares' on the AMITT grid) for practical use. This was to include advice on how to investigate each tactic - e.g. "This incident is using the pump priming tactic, what are the indicators you need to look for, what are the countermeasures you could use"", and how to counter it in the user's context.

1.2 Structure of this report

This report is structured around sections on AMITT design and use, and can be summarised as:

- This introduction section: history of AMITT's creation
- AMITT design and philosophy: why and how we built the frameworks, and the design choices we made
- AMITT component designs
- Using AMITT: Tools, technique and suggested uses
- Future work: notes and ideas for improving this work

This is a companion document to the AMITT TTP Guide and the master copy of the AMITT models, contained in github repository https://github.com/cogsec-collaborative/amitt.

This report is a living document. This is version 1.0.

1.3 Disinformation Defence

State actors, private influence operators, and grassroots groups are exploiting the openness and reach of the Internet to manipulate populations at a distance. This is an extension of a decades-long struggle for "hearts and minds" via propaganda, influence operations, and information warfare. Recent advances include computational propaganda: the use of algorithms, including machine learning and artificial intelligence, in online manipulation.

There are many definitions of misinformation, disinformation, incident etc. and teams dedicated to improving them. This report uses these working definitions:

- Disinformation is the deliberate promotion of false, misleading or misattributed information, usually designed to change the beliefs or actions of large numbers of people, as a tool to help meet an exterior goal.
- Misinformation is false or misleading information that's potentially harmful.

The structure and propagation patterns of misinformation attacks have many similarities to those seen in information security and computer hacking. Analyzing and building on similarities with information security frameworks gives defenders better ways to describe, identify and counter misinformation-based attacks.

1.4 AMITT

AMITT is a set of data standards and an open source knowledge base of both red team and blue team disinformation tactics and techniques. AMITT's intended users are disinformation responders; its purpose is to give them the ability to tactically respond to disinformation incidents, to plan defenses and countermoves, and to transfer information security principles to the disinformation sphere. It provides a common taxonomy for cognitive security offense and defence, a framework to rapidly share threat intelligence, and a conceptual tool for strengthening disinformation defences through red teaming, risk analysis, replays and simulations.

AMITT consists of blue team (defence) and red team (attack) models, and a repository of descriptions, mitigations and examples. To create AMITT, we placed misinformation components into a framework based on standards (including ATT&CK and STIX) commonly used to describe information security incidents. AMITT frameworks are designed to fit the same toolsets and use cases that STIX and ATT&CK are designed for.

1.5 Acknowledgements

AMITT was originally developed in 2019 by the Credibility Coalition's Misinfosec Working Group (MisinfosecWG), with inputs from the misinfosec community including experts who generously gave up a day of their time to workshop the first framework (the Atlanta workshop in 2019), then 2 days of their time to workshop potential counters (the DC workshop in November 2019).

The MisinfosecWG brainstormed, collated and devised new ways to counter or mitigate online manipulation, focusing on manipulation through disinformation and its known and potential countermeasures and mitigations. Our intent was always to give responders the ability to transfer other information security principles to the misinformation sphere, and to plan defenses and countermoves. For instance, we started the disinformation countermeasures workshop in Washington DC, with two main goals:

- Create the first version of a disinformation "Blue Team" playbook. For defenders, information security people and organizations, this will be a set of responses to misinformation attacks—the networks, the response types, the frameworks, and examples.
- Define how to support an operational global MisinfoSec_ISAO network. For potential response center participants and leaders, this will be the process, methods and understanding needed to connect, including suggesting partners, collaborators and funders.

MisinfosecWG was a short-term project to create information security-inspired standards for sharing information about misinformation incidents and how to respond to them. It was succeeded by the CogSecCollab nonprofit, which maintains the AMITT standards, with Sara-Jayne Terp and Pablo Breuer acting as design authorities.

CogSecCollab extended the original AMITT work, adding disinformation tools to infosec incident response and information sharing tools including MISP and STIX, and trialling the use of AMITT in its prototype disinformation Security Operations Centers (SOCs), including the CTI League's disinformation team, and with organisations including NATO, the EU and disinformation teams from several countries.

CogSecCollab has also been in discussions with MITRE about the MITRE team taking on AMITT alongside the ATT&CK model which inspired AMITT's design.

It takes a village, and we have many people to thank for their contributions to AMITT. Thank you all. We hope, with the new work, that we've done you proud.

2 AMITT Toolset Design and Philosophy

The AMITT toolset was created from a need for a common language for disinformation - at its creation time, our community included media, academics, infosec professionals, data scientists, government and people from other disciplines who all had different words for disinformation concepts and objects. AMITT tools should ideally provide a way for people from different fields to talk about misinformation incidents without confusion.

This section covers why and how we built the AMITT toolset, how its models connect to each other, and the design choices we made in their creation.

2.1 Disinformation as an Ecosystem

Our first move, back in 2016, was to talk about disinformation not as an isolated “fake news” problem, but as an ecosystem in which multiple actors with different motives (geopolitics, power, money, attention) interacted with misinformation and information flows, stories, beliefs, communities and individuals, websites, media, platforms and algorithms.

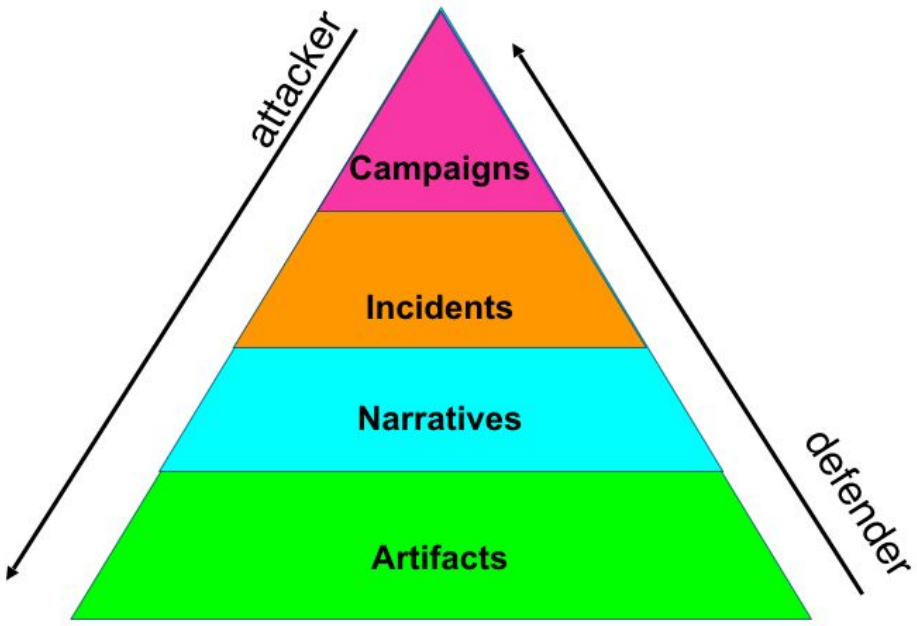
That was a lot of moving parts, and a lot of data, so we looked at other entities that were analysing ways that online and community beliefs and emotions could be changed, or analysing attacks on flows of information across the internet. These existing communities included social science, online marketing, adtech (online advertising technology) crisis data mapping, information security and data science. Our team all came from different directions on this, and all had different words and models for the same concepts, so in 2018, we formed two groups to connect them, and create a common language to talk about disinformation.

2.2 Connecting Defence Actors

When we started, we knew that our best chance of creating good disinformation defences meant connecting together people from very different worlds:

- The information operations specialists who spent their days analysing “conflict short of war” - military techniques like psychological operations (“psyops”), and other power moves between nationstates
- Data scientists, who analysed sets of objects and flows of information across the internet using techniques like machine learning and social network analysis to pick apart patterns of accounts, text, hashtags, urls, groups, and the relationships between them all.
- Social scientists and psychologists who studied human cognitive vulnerabilities, group dynamics, and the flow and effect of narratives on beliefs and emotions.
- Information security (infosec) experts, who had already built tools, techniques and processes to protect information held in very similar topologies, which instead of being communities of people were networks of machines.

Our first model, the disinformation pyramid, was built to help these groups talk to each other.



Disinformation Pyramid

Here we're looking at the different views that creators of misinformation ('attackers') and the people trying to counter them ('defenders') have (a third group involved, the targets of the misinformation, 'populations', aren't part of this diagram).

- Disinformation creators often persist in the ecosystem, focusing on one or more longer-term objectives (e.g. destabilize French politics, or reduce vaccination rates in target countries). We called these longer-term objectives "**campaigns**"; Clint Watts labelled these longer-term actors "advanced persistent manipulators" (APMs), mirroring the infosec term "advanced persistent threat". Many APMs are nation-state actors, using disinformation to attack other nations: this is the pyramid level that many of the information operations specialists were working at.
- **Incidents** are shorter-term sets of disinformation activity, often around a specific topic or event (e.g. Macrongate). These bursts of activity might be triggered by an event or opportunity to make money (there are many opportunists pushing misinformation), or they might be the result of a team of people working towards a desired effect: a change in beliefs or emotions relative to a specific person, group, object, concept, or event; or a weakening of an opposing group, belief etc by creating chaos and confusion. Campaigns typically contain multiple incidents, sometimes happening at the same time. Information security experts recognised this level of attack and mitigation as similar to the work they did deterring and mitigating attacks on information systems.
- **Narratives** are the stories that we base our beliefs on: "identity narratives" about who we are, "in-group" and "out-group" narratives about the groups that we do and don't belong to, and other narratives about what's happening in the world around us. Most incidents

use and rest on narratives, and we found ourselves tracking and talking about these as a useful abstraction of all the artifacts we collected for each incident. Narratives scientists fit into this layer of the pyramid, and it was a useful level to bring in social scientists and psychologists.

- **Artifacts** are the messages, images, accounts, relationships, and groups that a disinformation actor uses to create narratives and incidents. Artifacts are visible in each incident, often in large volumes, and are the disinformation layer that data scientists and other data specialists usually worked on.

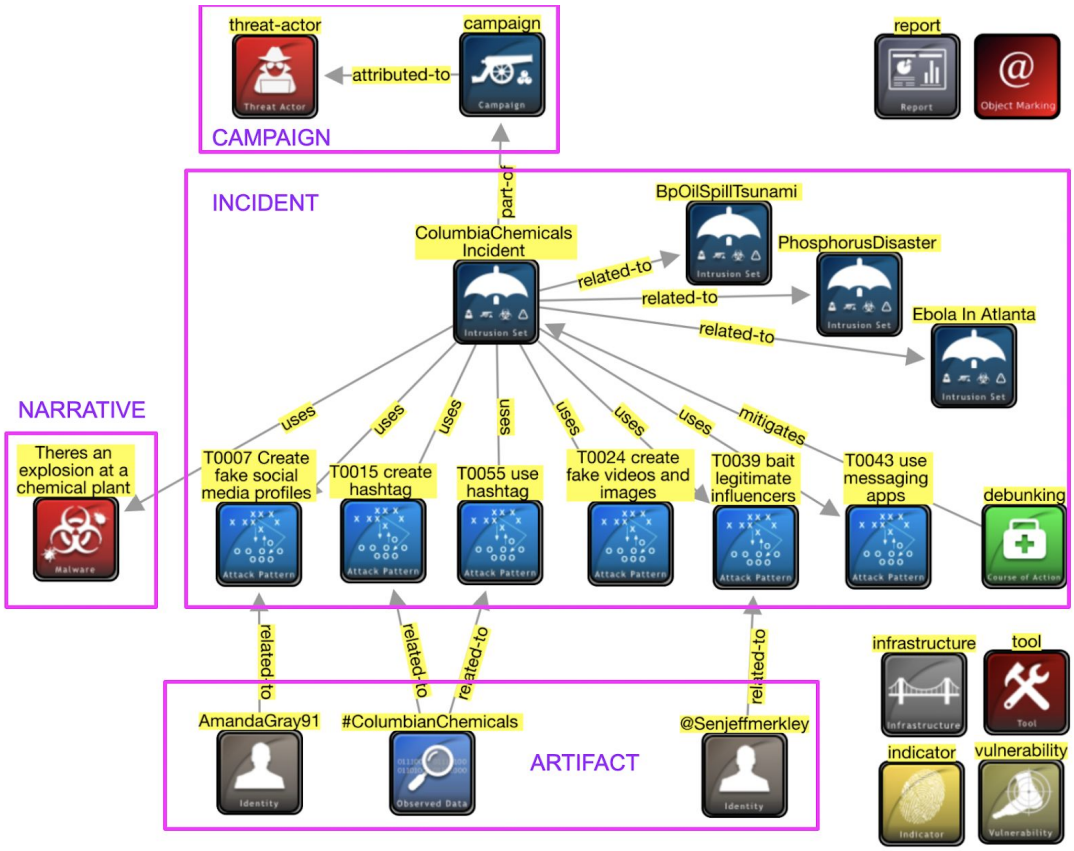
While the attacker sees the whole of the pyramid from the top down, the defender usually sees it from the bottom up, working back from artifacts to understand incidents and campaigns, unless they're lucky enough to have good insider information or intelligence, or have kept databases of information in forms that can be used to anticipate and compare artifacts, narratives etc to earlier work.

When we drew this pyramid in 2018, most of the misinformation tracking work that we saw was at the artifact level, with some work on the narrative (story) level, with Pablo coining the phrase that we were "admiring the problem" and needed to move to defence. Today (2020), all levels are investigated and connected, and the wider conversation has moved from tracking to defence and mitigation. We've also started to see human-readable reports on disinformation events that follow the layer model structure - starting with the wider context including campaigns, then an incident description including techniques used, then a list of narratives, and artifacts of interest at the end.

2.3 Component-based disinformation models

A useful view of a disinformation incident is as a collection of the objects seen within it, and the relationships between them. Many disinformation researchers already organise their information this way (as do the OSINT, intelligence and journalism-inspired research that much of this work is based on), with some of our earlier collaborators going as far as building "murder walls" to track groups and incidents. These are formalised as sociotechnical systems models - models of the complex networks of interacting communities, accounts and technologies that make up a disinformation incident or campaign.

The infosec community already has a data standard for this, STIX https://oasis-open.github.io/cti-documentation/, which also comes with a standard, TAXII, for how to share STIX data across systems. MisinfosecWG created a disinformation version of STIX, mapping its existing object types for disinformation use, and adding two new STIX object types: incident and narrative, because the existing objects, intrusion set and malware didn't quite fit what was needed for them.



STIX graph for the ColumbiaChemicals incident

A STIX graph for the Columbia Chemicals incident (a very short-term 2014 incident in the USA) is shown above, with the disinformation pyramid layers (campaign, incident, narrative, artifact) overlaid on it. This helps with thinking about relationships between disinformation layers: a disinformation incident usually belongs to one campaign (although there were many crossover campaigns in 2020, e.g. covid5g, where it was difficult to determine this), but multiple incidents can use the same narratives and artifacts.

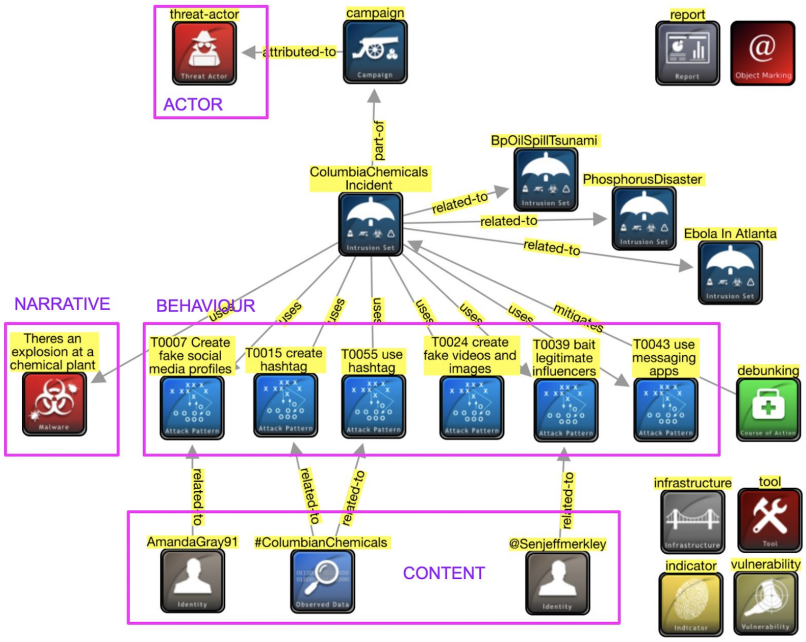
There are other component-based disinformation models, notably Camille Francois' ABC "Actor, Behaviour, Content" model and its extension, ABCDE ("Actor, Behaviour, Content, Degree, Effect"), which adds risk assessment components to assessing an incident.

Actor	*What kinds of actors are involved?* This question can help establish, for example, whether the case involves a foreign state actor
Behavior	*What activities are exhibited?* This inquiry can help establish, for instance, evidence of coordination and inauthenticity
Content	*What kinds of content are being created and distributed?* This line of questioning can help establish, for example, whether the information being deployed is deceptive.

Degree	*What is the overall impact of the case and whom does it affect?* This question can help establish the actual harms and severity of the case.
Effect	*What is the overall impact of the case and whom does it affect?* This question can help establish the actual harms and severity of the case.

ABCDE framework for disinformation [Pamment20]

ABC model components and narrative objects are shown in the ColumbiaChemicals diagram below - of note is that these model the disinformation creators' ABC, not the disinformation defenders (e.g. the "debunking" object is outside the Behaviour box).



ABC model components in the ColumbiaChemicals Stix graph

Building a disinformation model based on STIX allows analysts to share and compare information about threat actors, narratives, TTPs, artifacts and other objects in each incident and campaign, using the tools already built for STIX. It also allows disinformation data to flow through the same systems as information security data, making description and countering of hybrid (combinations of disinformation and other infosec methods) easier.

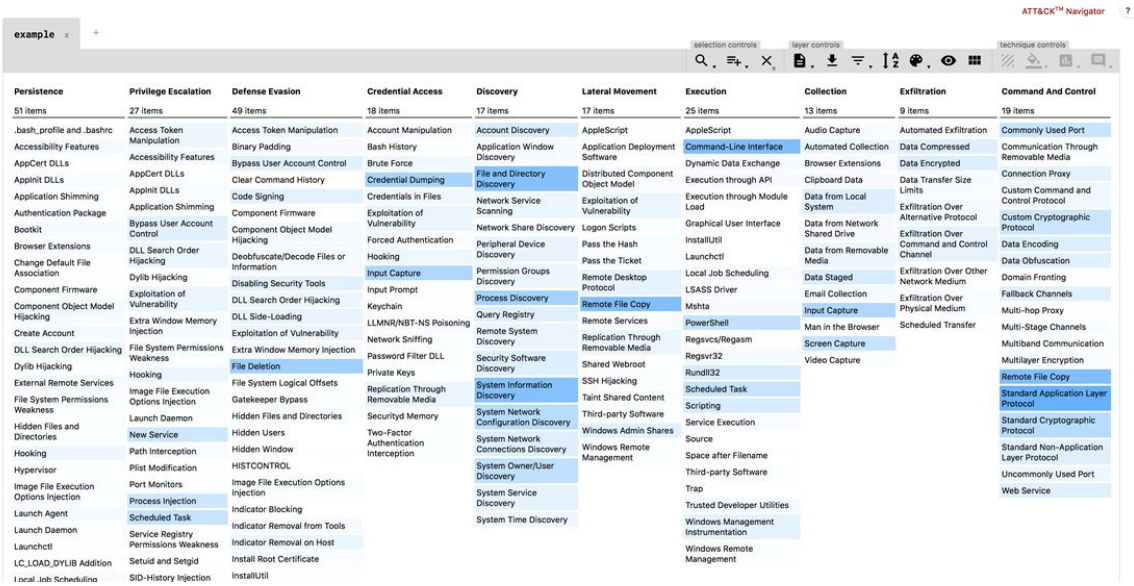
2.4 Behaviour-based Disinformation Models

The infosec community has multiple models that capture the behaviours of incident creators and responders. Several of these models, including MITRE's ATT&CK framework, focus on the techniques, tactics and procedures (TTPs) used by incident creators and responders, where TTPs are the blue ("attack pattern") and green ("course of action") boxes in the STIX diagrams

above. Most of MisinfosecWG's effort was on how to adapt these models, and the tools that use them, for disinformation response.

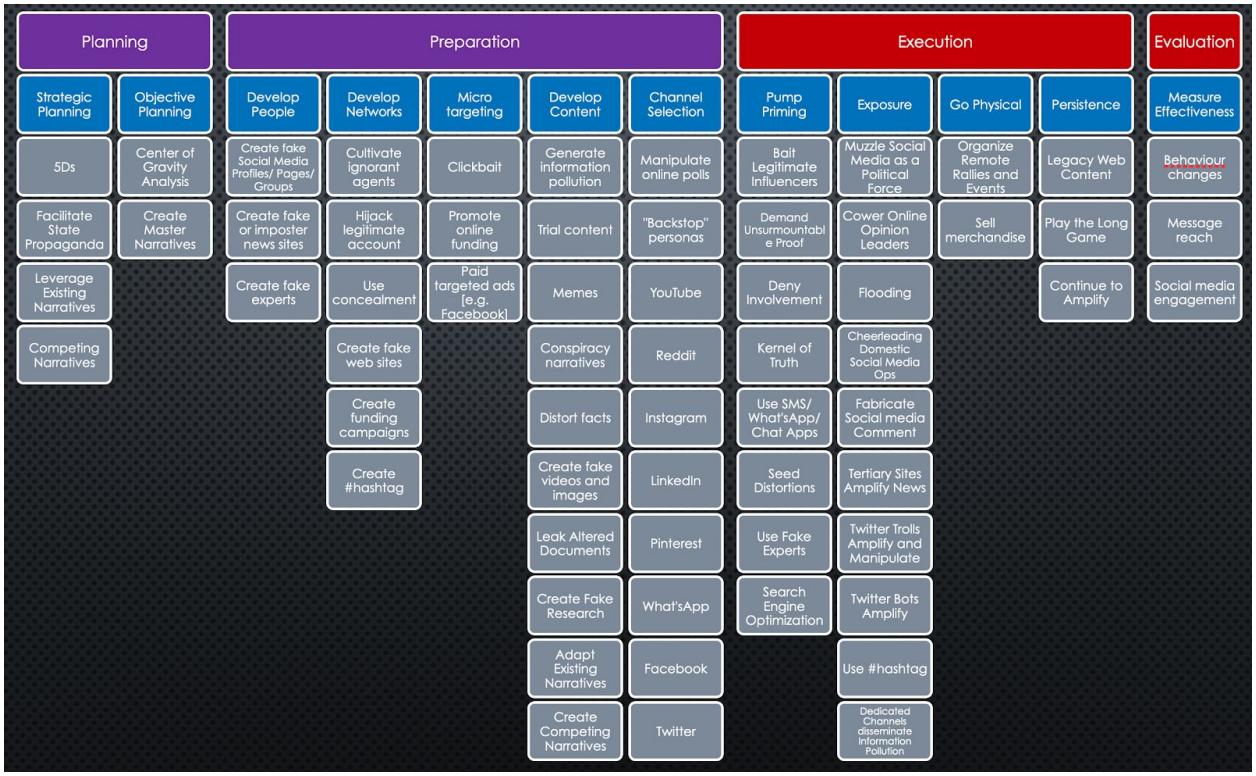
2.4.1 Disinformation Threat Models

In 2019, MisinfosecWG mined known disinformation incidents for incident creator behaviours, and looked for inspiration and frameworks for disinformation behaviour-based models. The group looked at behaviour-based models from information security, social network analysis, marketing, and adtech before settling on the Cyber Killchain and the ATT&CK model that's based on it, as a base representation for disinformation behaviours.



MITRE ATT&CK framework (Struse)

The model that MisinfosecWG created from this work is the AMITT Framework. The AMITT model (based on the ATT&CK framework) describes common disinformation TTPs across 12 stages of adversary activity, from strategic planning of each incident to evaluating its effectiveness and lessons learned from the deployment, as a feed into later incident plans.



AMITT Framework

The AMITT framework has three main component types:

- Phases (e.g. “Planning”)
- Tactic Stages (e.g. “Strategic Planning”): the set of top-level adversary goals that are needed to complete a successful attack.
- Tactics, techniques and procedures (TTPs, e.g. “5Ds”): the means by which incident creators meet each tactic goal.

With a behaviour-based framework, we can start to record and recall previous countermeasures to reused techniques, and find and exploit weaknesses and gaps in the adversary’s operations, in the same way we exploit adversary weaknesses in gaps in other situation pictures, including those in cybersecurity.

2.4.2 Disinformation Response Models

TTPs that model the behaviour of disinformation creators are one half of the behaviour-based models that we need for disinformation defence. In late 2019, MisinfosecWG extended its work to model the countermeasure and mitigation actions available to disinformation defenders.

Information security already has models for this: the course of action objects seen in STIX above. These are usually shown in a “Courses of Action Matrix” - a grid where tactic stages are plotted against different categories of countermeasure.

Table 1: Courses of Action Matrix

Phase	Detect	Deny	Disrupt	Degrade	Deceive	Destroy
Reconnaissance	Web analytics	Firewall ACL				
Weaponization	NIDS	NIPS				
Delivery	Vigilant user	Proxy filter	In-line AV	Queuing		
Exploitation	HIDS	Patch	DEP			
Installation	HIDS	"chroot" jail	AV			
C2	NIDS	Firewall ACL	NIPS	Tarpit	DNS redirect	
Actions on Objectives	Audit log			Quality of Service	Honeypot	

Cyber Killchain Courses of Action Matrix

A courses of action matrix for the cyber killchain (the model we based AMITT on) is shown above. Down the left side we have the seven cyber killchain tactic stages. Along the top we have six types of countermeasure or mitigation effect. Each grid square contains suggested actions that could create that effect on that tactic stage.

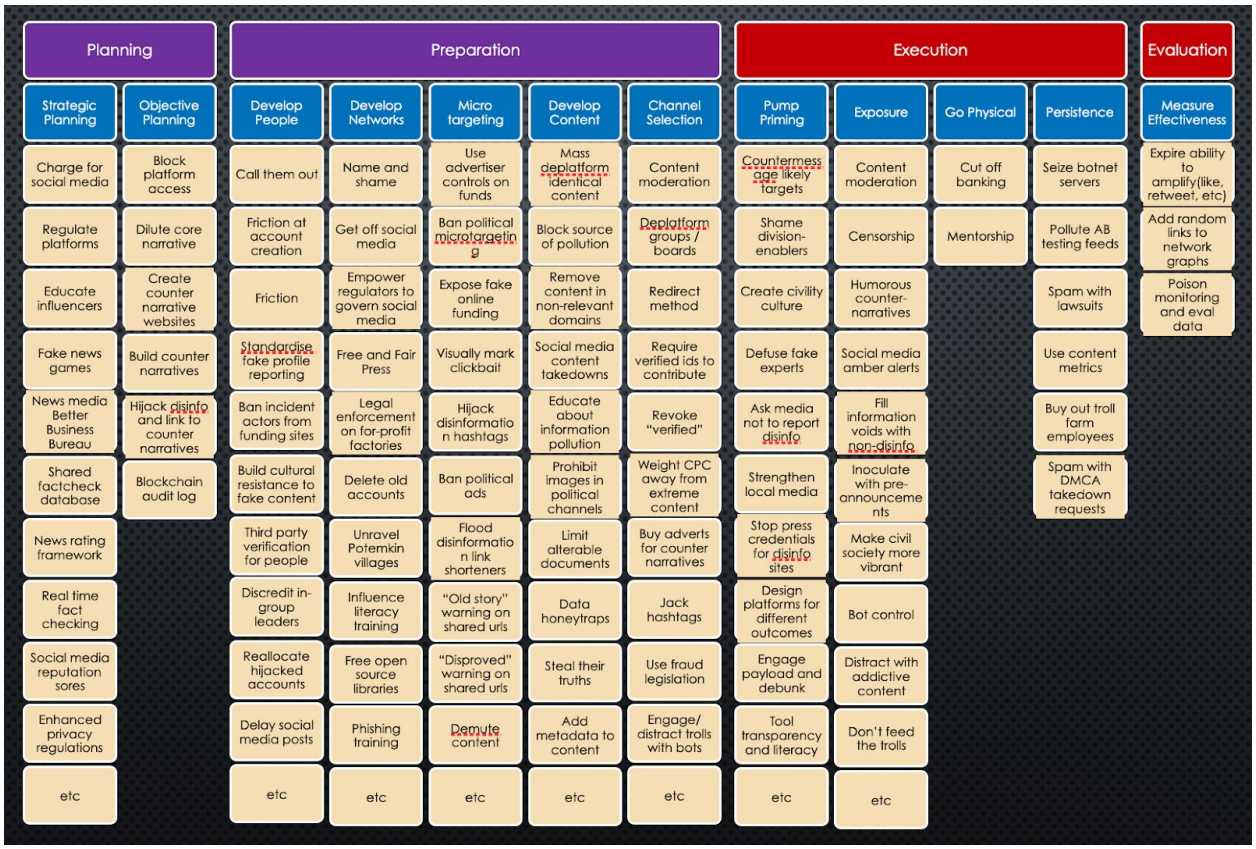
MisinfosecWG examined the disinformation solution space, considering the tools and techniques that existed and might be needed in it, then ran a Courses of Action generating exercise for the AMITT tactic stages, producing countermeasures and mitigations organised by countermeasure type, AMITT tactic stage and AMITT TTP. This formed a labelled list of AMITT disinformation creator TTPs that the CogSecCollab team extended to include the resources needed to deploy each countermeasure, and example playbooks for several counters.

	D2 Deny	D3 Disrupt	D4 Degrade	D5 Deceive	D6 Destroy	D7 Deter	TOTALS
TA01 Strategic Planning	11	6	7			4	28
TA02 Objective Planning		5					5
TA03 Develop People	10	7	1	1	1	1	21
TA04 Develop Networks	11	3	3		1		18
TA05 Microtargeting	2	5					7
TA06 Develop Content	13	8	5	2		5	33
TA07 Channel Selection	7	7	3	1			18
TA08 Pump Priming	7	3	2			3	15

TA09 Exposure	3	14	2				19
TA10 Go Physical	1					1	2
TA11 Persistence	1	6	6				13
TA12 Measure Effectiveness		1	2				3

AMITT counter TTP counts for AMITT tactic stages and counter types

This exercise produced more than 100 AMITT countermeasure TTPs that are now listed alongside the AMITT incident creator TTPs that they mitigate or counter.



AMITT countermeasures TTP diagram

The AMITT countermeasures TTP diagram is currently larger and messier than the AMITT incident TTPs diagram, as we work to clean it up and place techniques into the right stages.

2.4.3 Multiplayer game models

With the AMITT STIX, AMITT framework TTPs and AMITT countermeasure TTPs in place, it's possible to start modelling disinformation ecosystems as simulations or games in which multiple players compete for limited resources including narratives, attention and time. Threet designed models that focussed on resources, so multi-player, multi-viewpoint games and simulations could be designed using the existing AMITT TTPs and objects.

Another multi-player view of the disinformation solution space is as a human space, in which narratives compete for dominance (e.g “narrative warfare”). Human communication is generally at the level of stories, or narration: we tell each other stories about the world, as sentences, image sequences, or memes. Narratives are the stories that each person and community bases their sense of self, their belonging to different groups (“in-groups”), and exclusion of others (“out-groups”) on. Narratives are typically personal, emotionally-charged, deeply-entrenched and difficult to shift directly. In this space, it becomes important to track narratives and their components (e.g. memes, stories, sentiments) and disrupt them not by countering them directly with ‘facts’, but with ‘information aikido’: it’s easier to redirect an angry mob to a different house than it is to disband them. Narrative warfare is a growing field, and its techniques are a useful component in countering disinformation. Using Natural Language Processing techniques like topic modelling and gisting to track narratives from disinformation actors, and highlighting narratives to potential target audiences have also proved useful. AMITT models don’t explicitly include narrative warfare or machine learning models, although these have been built and studied independently by the AMITT teams.

2.5 Work in Progress

2.5.1 Disinformation Risk Modelling

Disinformation is a form of digital harm, alongside hate speech, cyber bullying, fraud, spam and other activities that potentially damage individuals, groups etc. digital harms can be managed as risks, where a risk is defined as a combination of severity, likelihood and target. SJ is working separately on disinformation risk models - these are useful in triaging (deciding which incidents to put response resources onto) misinformation, disinformation and threat actors.

In 2020, CogSecCollab used basic risk models to triage incidents coming into the CTI League and other deployments. These could be extended using the “DE” part of the ABCDE model, to give risk assessment and triage at other levels of the disinformation pyramid.

2.5.2 Disinformation Taxonomies

AMITT object types are not sufficient to completely describe a disinformation incident. AMITT STIX is missing categories for each of its object types.

For instance, AMITT STIX contains “threat actor”, but doesn’t have a set of labels for possible types of threat actor - geopolitically motivated, financially motivated etc, to make it easier for a user or information recipient to determine if a new actor is of interest to them.

Existing taxonomies of disinformation object types include DFRlab’s dichotomies of disinformation, which are designed for strategic analysis of disinformation actors, campaigns and incidents. In 2019, CogSecCollab worked with NATO to produce a taxonomy based on DFRlab’s taxonomy, but better suited for fast-paced tactical use.

2.5.3 Agile, and the limits of standards-based approaches

At this stage, older infosec people are probably shaking their heads and muttering something about stamp collecting and bingo cards. We get that. We know that defending against a truly agile adversary isn't a game of lookup, and as fast as we design and build counters, our counterparts will build counters to the counters, new techniques, new adaptations of existing techniques etc. Adversary tactics are moving quickly in this arena (for instance, the types of tool changes already seen in the related field of MLsec), so tools and counter tactics are likely to change but the basic problems won't.

But that's only part of the game. Most of the time people get lazy, or get into a rut — they reuse techniques and tools, or it's too expensive to keep moving. It makes sense to build descriptions like this that we can adapt over time. It also helps us spot when we're outside the frame.

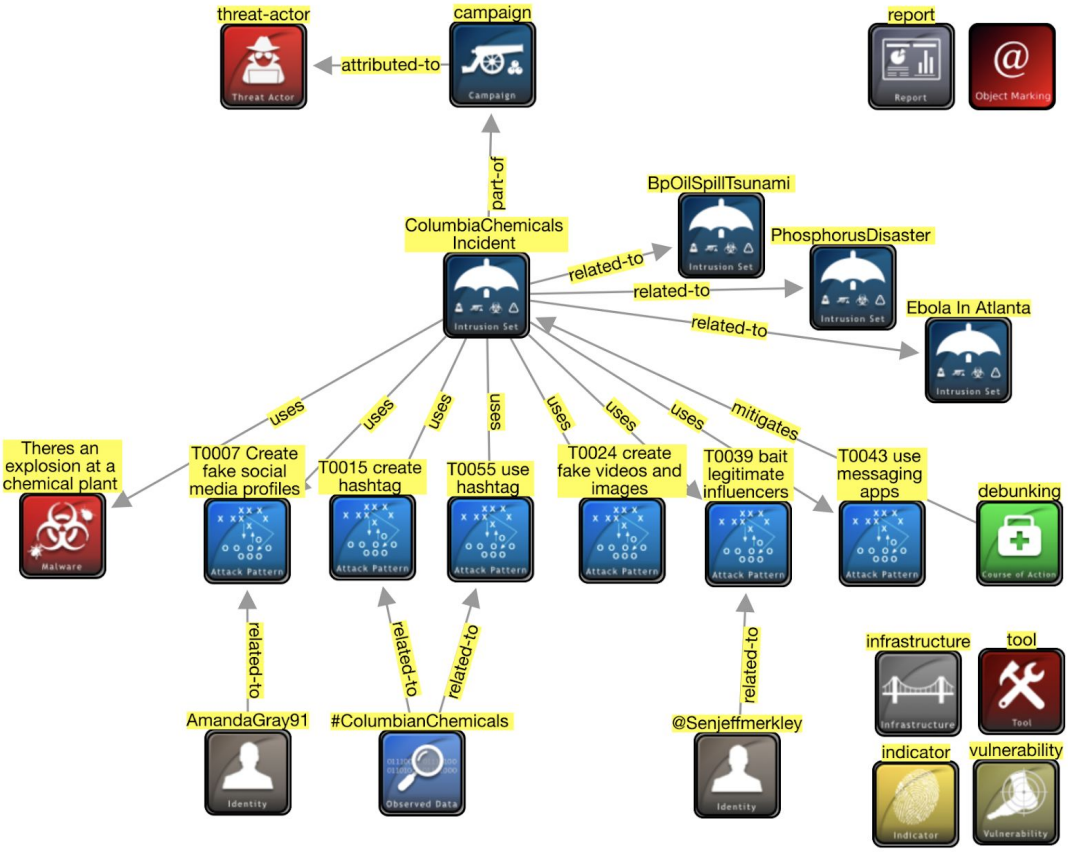
There is no one, magic, response to misinformation. Misinformation mitigation, like disease control, is a whole-system response. All the tools mentioned above are intended for use by threat intelligence teams, often working in near-real-time from Security Operations Centers and their equivalents.

Sometimes you just respond, but it helps to do this from a position of knowledge, shared communication and respect for the potential risks to actors, organisations, narratives and other components of the information ecosystem we're working in. MisinfosecWG looked at Adam Shostack's slides on threat modelling in 2019, and specifically at the differences between slower "waterfall, V" threat models (STRIDE, kill chain etc), and faster-reacting "agile" and lean threat models, where agile is rapidly iterating over solutions in a known problem space, and lean is iterations on both the problem and solution spaces. This is one of the considerations when designing tactical disinformation response: we still need the slower, deliberative work that gives labels and lists defences and counters for common threats (the "phishing" etc equivalents of cognitive security), but we also need that rapid response to things previously unseen that keeps white-hat hackers glued to their screens for hours. There's more about this in CogSecCollab's writings on creating and operating disinformation Security Operations Centres.

2.6 Further Reading

- WWW 2019 AMITT paper; summary of AMITT WWW paper

3 AMITT STIX Design and Philosophy



STIX graph for the Columbia Chemicals incident

STIX is a data standard used to share information between threat intelligence organisations like ISACs. It's a rich language that describes threat objects and the relationships between them, is extensible, used by existing threat intelligence sharing communities \(ISACs, ISAOs etc\) so we'd be patching into an existing sharing system. It's also supported by and integrates well with existing community-supported, open-source tools. STIX translates well for disinformation use.

Disinformation STIX	Description	Level	Infosec STIX
Report	communication to other responders	Communication	Report
Campaign	Longer attacks (Russia's interference in the 2016 US elections is a "campaign")	Strategy	Campaign

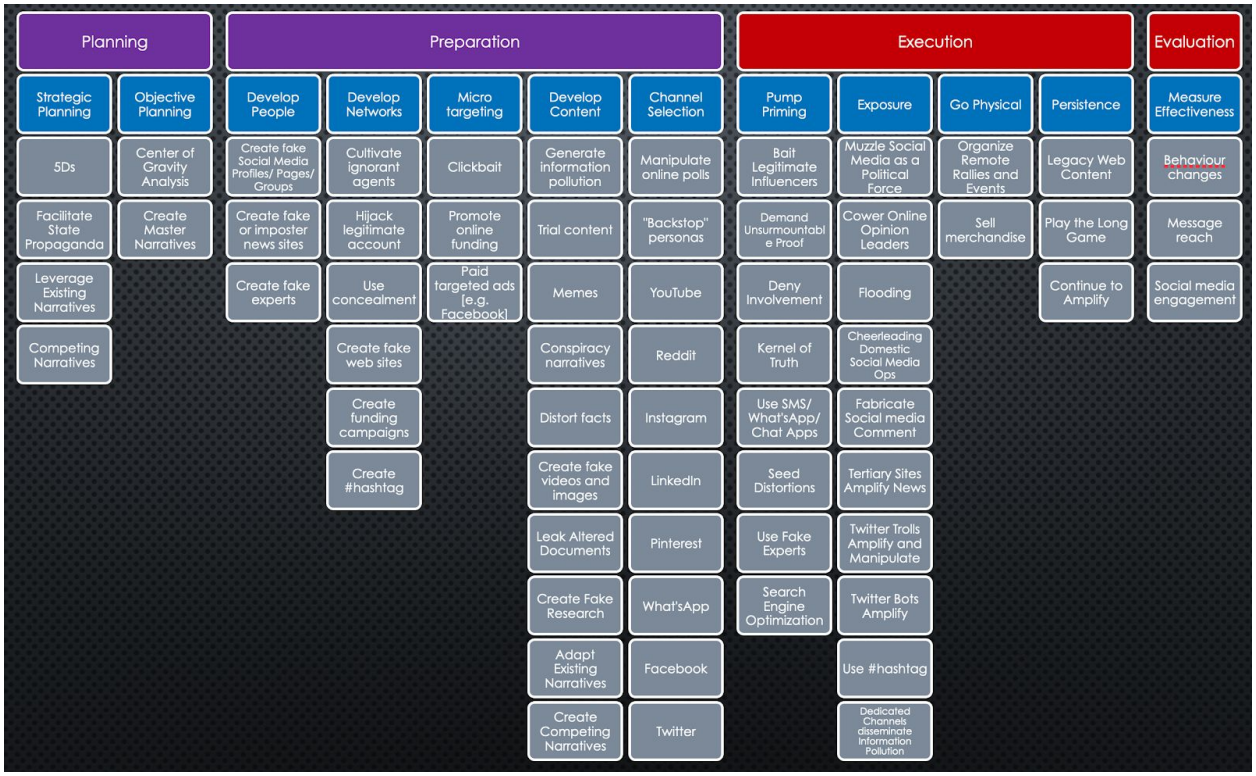
Incident	Shorter-duration attacks, often part of a campaign	Strategy	Intrusion Set
Course of Action	Response	Strategy	Course of Action
Identity	Actor (individual, group, organisation etc): creator, responder, target, useful idiot etc.	Strategy	Identity
Threat actor	Incident creator	Strategy	Threat Actor
Attack pattern	Technique used in incident (see framework for examples)	TTP	Attack pattern
Narrative	Malicious narrative (story, meme)	TTP	Malware
Tool	bot software, APIs, marketing tools	TTP	Tool
Observed Data	artefacts like messages, user accounts, etc	Artefact	Observed Data
Indicator	posting rates, follow rates etc	Artefact	Indicator
Vulnerability	Cognitive biases, community structural weakness etc	Vulnerability	Vulnerability

Mappings Between infosec STIX and cogsec STIX

We added two objects to STIX for disinformation: incident, and narrative, and didn't need to change anything else. We use custom objects to represent these fields and be OpenCTI compliant.

AMITT is now available as a STIX 2.0 bundle, from https://github.com/cogsec-collaborative/amitt_cti . When STIX 2.1 delivers an incident object we'll migrate to that.

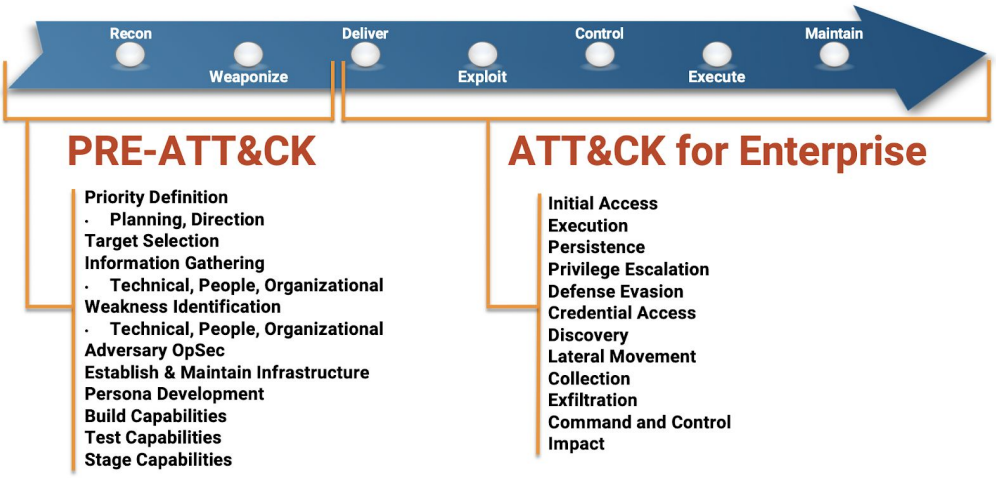
4 AMITT Framework Design and Philosophy



AMITT Framework

The AMITT framework was created from a need to describe disinformation behaviours in consistent, concise ways that could allow rapid sharing of information across responding groups.

Seeding the Model



Top: Cyber Killchain stages, Bottom: ATT&CK framework stages (MITRE intro to ATT&CK)

MisinfosecWG mapped the other main models it was considering for the AMITT framework onto the cyber killchain, to ensure it missed as little as possible from them.

Marketing 1	Marketing 2	Cyber Killchain	Psyops phases	Justice Department	New York Times
		RECON	1. Planning	Research (target environment)	
Market research	Market research		2. Target audience analysis		Find the cracks
Campaign design	Campaign design		3. Series development		Seed distortion
		WEAPONIZE		Position (infrastructure + networks)	Wrap narratives in kernels of truth
Content production	Content production		4. Product development and design, 5. Approval	Produce (content)	
Awareness	Exposure	DELIVER	6. Production, distribution, dissemination	Publish (content dissemination)	Build audiences
	Discovery				
Interest/Consideration	Consideration				
Conversion/Purchase	Customer relationship	EXPLOIT			
		CONTROL			
		EXECUTE			
Loyalty/Retention	Retention	MAINTAIN			
Advocacy				Amplify (media saturation)	Cultivate "useful idiots"
					Deny involvement
			7. Evaluation	Calibrate (assessment +retooling)	Play the long game

Comparison between cyber killchain, marketing, psyops and other potential models

4.2 Organising the AMITT Framework

AMITT's phases are grouped into activities typically performed before a disinformation incident become publicly visible, and those after incident artifacts are widely visible online. The phases before public visibility are termed "left of boom"; those after are "right of boom" (this is an old explosive disposal term used in some infosec models).

Like ATT&CK, AMITT's tactic stages are listed sequentially from left to right - the further left that a tactic is on the AMITT diagram, the earlier that it's likely to be met by an incident creator. In AMITT, tactics are also grouped into four phases: planning, preparation, execution and

evaluation; phases are used to evaluate things like the potential for both attacker and defender automations.

Each AMITT TTP description includes examples of its use, defender TTPs that could be used to counter or mitigate it, and indicators that could be used to detect it.

Sub-techniques are lower-level, very specific techniques. Sub-techniques aren't shown on the main AMITT framework diagram.

Tactic Phases

Tactic stage	**Threat actor is trying to…**	**Techniques**	**KillChain Phase**
Strategic planning	Define the desired strategic end state of the incident.	4	Recon
Objective Planning	Create clear, measurable, and achievable tactical task objectives for the incident.	2	Recon
Develop people	Develop online and offline users and agents, including automated personas	3	Weaponize
Develop networks	Develop online and offline communities and transmission methods	6	Weaponize
Micro targeting	Target very specific populations of people	3	Weaponize
Develop content	Create and acquire content used in incident	10	Weaponize
Channel selection	Set up specific target, delivery, amplification and manipulation channels for incident	10	Weaponize
Pump priming	Release content on a targeted small scale, prior to general release, including releasing seed narratives	8	Deliver
Exposure	Release content to general public or push to larger population	10	Execute
Go physical	Move incident into physical world	2	Execute
Persistence	Keep incident 'alive', beyond the incident creators' efforts	3	Maintain
Measure Effectiveness	Measure effectiveness of incident, for use in planning future events	3	Maintain

4.3 Further Reading

ATT&CK models

- MITRE, "ATTACK_Design_and_Philosophy", 2020

- MITRE, “Getting started with ATT&CK”, October 2019

AMITT

- SJ Terp, “Misinformation has stages”, Misinfocon 2019

5 AMITT Countermeasures Design and Philosophy

The AMITT countermeasures framework was created from a need to move from "admiring the problem", to actively responding to and mitigating for disinformation in as close to real time as sensibly possible.

This section looks at existing and potential disinformation countermeasures and mitigations. It's part of a series of work on how information security principles and practices can be used to improve our understanding of and responses to disinformation campaigns and incidents.

5.1 Finding Countermeasures

5.1.1 Introduction

But right now, it's still part of the "admiring the problem" collection of misinformation tools -to be truly useful, AMITT needs to contain not just the breakdown of what the blue team thinks the red team is doing, but also what the blue team might be able to do about it. Colloquially speaking, we're talking about countermeasures here.

There are several ways to go about finding countermeasures to any action:

- Look at counters that already exist. We've logged a few already in the AMITT repo, against specific techniques — for example, we listed a set of counters from the Macron election team as part of incident I00022.
- Look at AMITT's parent models - the ATT&CK framework, the psyops model, marketing models etc - and see how they modelled and described counters (e.g look at the mitigations for ATT&CK T1193 Spear phishing).
- Pick a specific tactic, technique or procedure and brainstorm how to counter it — the MisinfosecWG did this as part of their Atlanta retreat, describing potential new counters for two of the techniques on the AMITT framework.
- Wargame red v blue in a 'safe' environment, and capture the counters that people start using. The Rootzbook exercise that Win and Aaron ran at Defcon AI Village was a good start on this, and holds promise as a training and learning environment.
- Run a machine learning algorithm to generate random countermeasures until one starts looking more sensible/effective than the others. Well, perhaps not, but there's likely to be some measure of automation in counters eventually…

MisinfosecWG mapped out misinformation responses, e.g.

- At the technique level — T0025 leak altered documents was countered in France during the Macron election.
- At the tactic level — we can create a courses of action matrix that lists ways to detect, deny, disrupt, degrade, deceive or destroy activities in each tactic stage.

- At the procedure level — we can look at sequences of responses that may be more effective than individual responses in isolation.

5.1.2 Searching for Countermeasures

Searching for disinformation resources at the end of 2019 is much easier than in previous years. Major lists of projects, reports and groups that yielded existing countermeasures included

- Oxford Internet Institute's computational propaganda project's resource finder https://navigator.oii.ox.ac.uk/resources/?resource_filter%5Bsubject%5D%5B%5D=disinformation-counter-strategies#
- Rand.org's reports on disinformation (e.g. [Rand2740])
- Scott Yate's CCC lists of projects, and the Credibility Coalition's navigator

Many other groups (CMU etc) are creating their own lists, making this a great time to hunt for specific counters.

5.1.3 Known Countermeasures

There are many published "solutions" to disinformation attacks. While useful, it's foolish to consider any of these the "silver bullet" that solves a disinformation problem; they often address smaller pieces of an attack, or are intractable or don't scale. Disinformation campaigns are whole-system attacks: to solve them we need to look at whole-system solutions: this is more of a "thousand bullet" solution than a single-bullet one. Some components in the current counter landscape are:

- Detecting artificial amplification. Many disinformation campaigns rely on signal amplification, either through 'useful idiots' or by raising message visibility using non-human traffic ('bots' and 'botnets'). Databases of known online bad actors and state-sponsored actors, with data from pages and social media feeds from these actors have proven useful places to look for emerging narratives and links to new actors and artefacts. Tracking bots and botnets has become more difficult as adversaries adapt to detection techniques (both from disinformation detection but also from adjacent domains including mitigating advertising click fraud) and trade message reach for keeping valuable networks online, but there is still value in simple bot/botnet detection techniques including analysis of similarities across accounts linked by topic, hashtags, retweets, references etc, and time-series analysis to check for sleep/wake patterns, activity correlations etc, especially with adversaries new to this space.
- Detecting related artifacts. Disinformation campaigns rarely use one account, platform, account network or domain, and financially-motivated campaigns sometimes run sets of sites with wildly different topics or demographic/country targets. Most work on this isn't tool-based; it's digital forensics, tracking artifacts like tag IDs, domain registrations and reused/linked content across the internet using OSINT tools (Bellingcat and DigitalSherlocks both publish good examples of this work).

- Mitigating artificial amplification. Most current work on this is platform takedowns or "shadow-banning" of known bot, botnet, troll or other artificial amplification social media accounts. Related work includes removal of online advertising and product revenue from domains that are part of financially-motivated disinformation campaigns.
- Resilience against adversarial narratives. It's preferable to remove a disinformation campaign before it reaches the general population, but if it does, building resilience to disinformation campaigns in the form of awareness of techniques, critical reasoning skills etc is useful. Most population resilience counters are in the form of education - either at school level or through information campaigns like the US State Department's War on Pineapple posters. More active population resilience measures include the Baltic Elves volunteer groups posting disclaimers and counter-narratives to Russian disinformation in their countries.

Education is an important counter, but won't be enough on its own. Other counters that are likely to be trialled with it include:

- Tracking data providence to protect against context attacks (digitally sign media and metadata in a way that media includes the original URL in which it was published and private key is that of the original author/publisher)
- Forcing products altered by AI/ML to notify their users (e.g. there was an effort to force Google's very believable AI voice assistant to announce it was an AI before it could talk to customers)
- Requiring legitimate news media to label editorials as such
- Participating in the Cognitive Security Information Sharing and Analysis Organization (ISAO)
- Forcing paid political ads on the Internet to follow the same rules as paid political advertisements on television
- Baltic community models, e.g. Baltic "Elves" teamed with local media etc

Jonathan Stray's paper "Institutional Counter-disinformation Strategies in a Networked Democracy" is a good primer on counters available on a national level.

Table 1: Counter-disinformation strategies used by the three institutions in this paper, and their effectiveness and legitimacy in a democratic society.

Strategy	Used by	Effectiveness	Legitimacy
Refutation	EU Stratcom Facebook via fact-checkers	Works if consistent, but not all disinfo is about facts.	Generally legitimate to speak the truth, though people will disagree on what truth is.
Expose inauthenticity	EU Stratcom Facebook	Discredits the source, provides justification for further measures.	Content-neutrality is appealing. Important to preserve legitimate anonymity.
Alternative narratives	EU Stratcom China	Helps displace disinfo, inoculates against it if seen first.	Can itself be disinfo or distraction.
Algorithmic filter manipulation	Facebook China via 50c party	Media algorithms have huge effect on information exposure.	Platforms may abuse this power, users may game it.
Speech laws	Facebook enforces such laws China	Can be effective at targeting narrow categories of speech.	Broad laws against untruth are draconian.
Censorship	China	Effective when centralized media control is possible.	Generally conflicts with free speech.

"A taxonomy of tactics" [Stray19]

5.1.4 Countering AMITT components

Work on AMITT used existing information security models (e.g. cyber kill chain, ATT&CK) to model disinformation incidents as collections of tactics, techniques and procedures (TTPs). One way to look at counters is to look at that breakdown and find or devise responses and mitigations to each TTP. At the tactic level, this gives us a Courses of Action matrix (COA), with the tactic stages listed on one axis, and types of response - eg. (Deny, Disrupt, Degrade, Destroy) - on the other, At the technique level, this gives us a way to discuss mitigations for techniques (e.g. the use of botnets) that we see repeatedly in disinformation incidents.

This is one way to look at countermeasures and mitigations. It's a useful way to examine the space of possible actions, in the same way that a naval officer learns about 'standard' manoeuvres like the Crazy Ivan, and how to think about detecting and mitigating for them. Disinformation, like war, isn't a linear process: that there are techniques in play that work and are likely to be used is just the first level of understanding what could and might be done. Good

incident creators are also artists (yes, yes, there's a reason it's called "the Art of War"), understanding the basic techniques and constraints, and knowing how to adapt them into a flow of actions that becomes difficult to counter with a simple rulebook. These masters still need to know the basics though.

5.1.5 Workshopping Counters

Day 1

- Introduce what MisinfoSec_WG has done, why we've done it, and what we have to show. Introduce AMITT; review stages and techniques
- Workshop/hands on "Blue Team" to build the responses part of the framework
- Create 5-7 five-person multi-disciplinary teams each responsible for creating a collection of counters for up to 10 of the 54 identified techniques

Day 2

- Introduce ISAO concepts and how they connect to misinformation
- Workshop/hands on design of ISAO network support
- Workshop/hands on exercise testing responses and network concept together
- Wash-up and next steps

5.2 AMITT Countermeasure components

When organising countermeasures, there are a few questions to ask:

- What does this counter do? Is this a mitigation, and what does it do: does it stop a technique being effective, moderate its effect or do something like delay its effect whilst other measures are put in place?
- Who can do this? What skills and resources do they need to have a chance at success? What risks do they take in doing it and how can those be both explained and minimised?
- Has this been tried before? What happened that time? Are there side effects (both good and bad) to watch out for?
- Has this been used in combination with other counters? Could it be?
- What level is this counter at? Is it strategic, tactical or immediate?

Answering these questions meant adding appropriate labels and examples to each countermeasure. This subsection covers some of those labels.

5.2.1 Countermeasure types

The list of countermeasure types is a cut-down version of the US Military's Joint Publication 3-13, aka JP3-13 This descriptions of the list items appears on page I-9:

“Objectives:

Commanders use IO capabilities in both offensive and defensive operations simultaneously to accomplish the mission, increase their force effectiveness, and protect their organizations and systems. Fully integrating IO capabilities for offensive and defensive operations requires planners to treat IO as a single function. Commanders can use IO capabilities to accomplish the following:

1. *Destroy. To damage a system or entity so badly that it cannot perform any function or be restored to a usable condition without being entirely rebuilt.*
2. *Disrupt. To break or interrupt the flow of information.*
3. *Degrade. To reduce the effectiveness or efficiency of adversary C2 or communications systems, and information collection efforts or means. IO can also degrade the morale of a unit, reduce the target’s worth or value, or reduce the quality of adversary decisions and actions.*
4. *Deny. To prevent the adversary from accessing and using critical information, systems, and services.*
5. *Deceive. To cause a person to believe what is not true. MILDEC seeks to mislead adversary decision makers by manipulating their perception of reality.*
6. *Exploit. To gain access to adversary C2 systems to collect information or to plant false or misleading information.*
7. *Influence. To cause others to behave in a manner favorable to US forces.*
8. *Protect. To take action to guard against espionage or capture of sensitive equipment and information.*
9. *Detect. To discover or discern the existence, presence, or fact of an intrusion into information systems.*
10. *Restore. To bring information and information systems back to their original state.*
11. *Respond. To react quickly to an adversary’s or others’ IO attack or intrusion*

All IO capabilities may be employed in both offensive and defensive operations.”

Action types *exploit, influence, protect, restore and respond* weren’t viewed as immediately relevant to disinformation work.

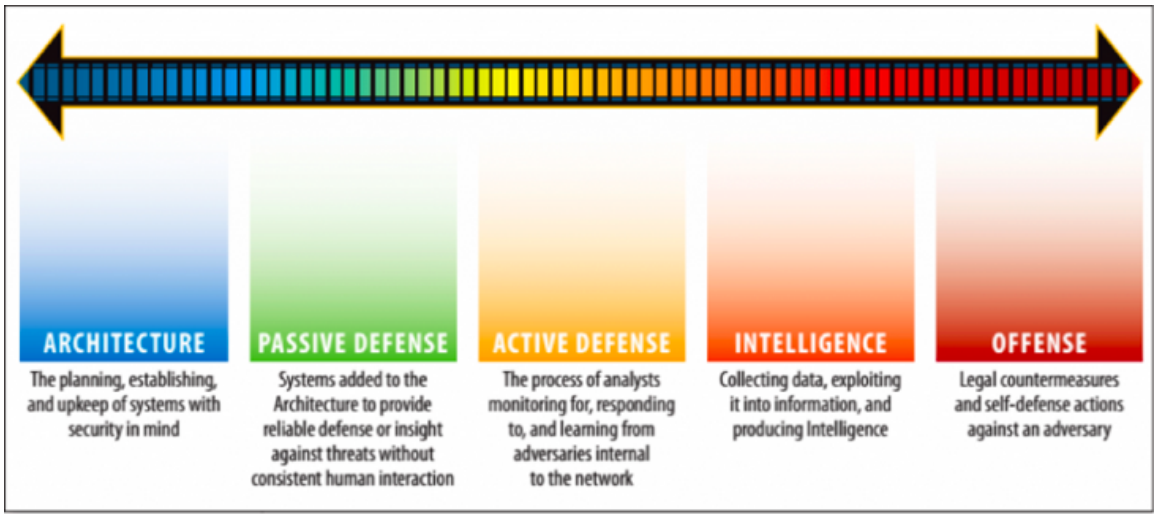
5.2.2 Response Actors

Describing actions is great, but actions only work if someone does them. There are many entities in the space of being affected by and analysing disinformation campaigns; not so many entities in the space of being able to, willing to, legally allowed to, or actively responding to disinformation. Entities who could respond include social media platforms, other organisations, civil society, media organisations, governments, militaries and individuals. There are also other stakeholders who could be persuaded or find it in their best interests to help reduce the prevalence of disinformation campaigns across societies.

- Social media platforms have control over their own software, and usually have control over the data moving through it, and the data available on and archived in it. They also have control over who can access that software and data - or rather, over which accounts can access it. Very few social media companies are owned by individuals now - they tend to be accountable to business stakeholders whose motivation is, generally, profit. This means that removing disinformation from systems is often in competition or conflict with other business priorities, or may require system adaptations or rebuilds that are too costly to justify against an uncosted, unquantified, unknown damage to society.
- Other online organizations include organizations like web hosts and DNS registrars, who could help with the removal of disinformation campaign websites.
- Civil society is that connector between the people trying to help counter disinformation campaigns and the people who are subjected to them. This is where people-centre approaches like education and reporting routes for microtargeted messages and advertisements are tried.
- Media has its own disinformation problems, despite its emphasis to itself on trying to find truth. Falling media budgets, longer/faster news cycles and wide access to information about breaking stories has left individual net journalists struggling to keep up and wade through streams of information, malinformation and disinformation around events. The counters here are two-way - both journalists helping counter disinformation with new practices (e.g. “rumour” pages during natural disasters), and in better training on content ingestion and dissemination practices.
- Governments can help primarily with the regulations that companies can use to justify moving disinformation measures above other line items in their business plans. The shadier parts of government can also help with more direct action tracking down and dissuading campaign creators and amplifiers.

5.2.3 Meta Techniques

There are legal restrictions in many countries on the types of counter response that different actors can perform: for example, in the United States, the Posse Comitatus Act limits offensive actions of US military on US territory, making the lists of potential actors fraught with questions like “yes, this group of responders could use this countermeasure, but is it legal and/or moral of them to do so?”. Circumventing Posse Comitatus by using the National Guard notwithstanding, one of the first actions in answering that across multiple countries is to label counter TTPs by whether they’re offensive or not.



SANS scale

Information security has a framework for this too: the SANS scale, as shown above. In many cases, this was too coarse grained a scale to help with determining who could potentially use a measure, so we also tagged counter TTPs with the rough type of action they were suggesting, as seen below.

Metatechnique	Description	SANS
metatechnique	Not direct counters, but fit the SANS architectural level of countering	architecture
cleaning	Clean unneeded resources (accounts etc) from the underlying system so they can't be used in disinformation	passive
data pollution	Add artefacts to the underlying system that deliberately confound disinformation monitoring	passive
daylight	Make disinformation objects, mechanisms, messaging etc visible	passive
diversion	Create alternative channels, messages etc in disinformation-prone systems	passive
resilience	Increase the resilience to disinformation of the end subjects or other parts of the underlying system	passive
scoring	Make scores available	passive
counter messaging	Create and distribute alternative messages to disinformation	active
dilution	Dilute disinformation artefacts and messaging with other content (kittens!)	active
friction	Slow down transmission or uptake of disinformation objects, messaging etc	active
reduce resources	Reduce the resources available to attackers	active
removal	Remove disinformation objects from the system	active
verification	Content	active
targeting	Target attackers	offense

AMITT Meta Technique categories

This provides a bridge between the disinformation types and the SANS scale.

5.3 Building AMITT-based Playbooks

A collection of countermeasures is nice, but it's not going to help someone who's facing an immediate active campaign or incident. They're going to need some form of "hey, this is happening, here are things you could try and what might happen" guides.

One of the things that reading through the counters spreadsheet surfaces is the sense of who is doing what to whom with which resources? For example - we have a lot of entries that look something like "tell x about y". Which is great, but that assumes that y can do something about x. After a while this starts to look like pieces of a stix graph itself - we have actors (or types of actor), artefacts and techniques in play, connecting to and relying on each other. Content takedowns, for instance: these can only happen if the people capable of doing the takedowns know about the content, and the people who know about the content tell them about it. We may also have a componentwise, piece-together set of responses to be built. To start with, mapping out who is doing what to whom with which resources, and which assumptions about actions and outcomes might go a long way in reducing our 200+ listed counters down to a manageable tactical set.

5.4 Further Reading

Must-reads on counters

- [Stray19] Jonathan Stray, "Institutional Counter-disinformation Strategies in a Networked Democracy", WWW 2019 (video)
- The war on pineapple: https://www.dhs.gov/sites/default/files/publications/19_0717_cisa_the-war-on-pineapple-understanding-foreign-interference-in-5-steps_0.pdf
- Chapter 7 of https://www.foreign.senate.gov/imo/media/doc/FinalRR.pdf

General references on counters

- https://ukraineelects.org/live-updates/page/4/
- https://navigator.oii.ox.ac.uk/resources/?resource_filter%5Bsubject%5D%5B%5D=disinformation-counter-strategies#
- https://www.climatechangecommunication.org/wp-content/uploads/2020/10/DebunkingHandbook2020.pdf

6 Multi-Player Game Models: design and philosophy

One potentially fruitful model of disinformation is as a game where multiple players on both red and blue teams compete and cooperate for resources, using the TTPs from the AMITT framework and AMITT counters models. In 2020, CogSecCollab ran weekly red team exercises, usually based on incidents the team was tracking or countering online. These exercises used the AMITT TTPs, meta-techniques and STIX objects, with realistic estimates of red and blue team resources, to anticipate new disinformation narratives and moves. These were used to help prepare mitigations and watches for future incidents, and draft a "Doctrine for countering disinformation".

Much of this work was on the operational level, using the DIMEFIL model for geopolitics and business, and TTPs to model manoeuvres in those spaces.

Critical resources included:

- Resources
    - Transmission media
    - Audience
    - Message generation (narratives?)
    - Manhours
    - Intelligence, Access, Capability
    - Credibility
    - "Money, money, and money" (Trivulzio)
- Message
- Credibility
- Access/Audience
- Temporal (timelines, deadlines)

This work used a single Centre of Gravity. The most critical resource we found was time, e.g. to delay, scope, front-run, etc.

Other recent work on AMITT and multi-player games plots the disinformation red team and blue team TTPs for an incident together, and tracks their connections and potential effects on each other.

6.1 Further Reading

Susan Young, Dave Aitel, "The Hackers Handbook"

7 Ways to Work with AMITT

AMITT is designed for rapid sharing of disinformation threat intelligence, over the same systems used by information security professionals, but that's not the only thing you can do with it. This section covers uses and tools, including:

- Offensive planning
- Defensive planning
  - Red teaming and planning
  - Intelligence analysis and tracking
  - Active defence and countering
- Sharing AMITT data
- Tools (including MISP) for disinformation tracking

We've used the AMITT framework to decompose different misinformation incidents into stages and techniques, so we can start looking for weak points in the ways that incidents are run, and in the ways that their component parts are created, used and put together. We've also used it to analyse what's possible in terms of algorithm use and other automations.

7.1 AMITT Trials and Implementations

In 2020, we used AMITT in live and test disinformation defence deployments.

7.1.1 AMITT MISP Implementations

AMITT was implemented, tested and used in two MISP instances:

- The CogSecCollab MISP instance, used for testing by both CogSecCollab and other groups trialling the AMITT framework.
- The Covid19 MISP instance, used by groups around the world to share threat intelligence about Covid19 information security issues.

The CTI League's Disinformation team, led by CogSecCollab team members, worked with the Covid19 instance, adapting tools, processes and models to fit a team handling large volumes of information at rapid speed. Innovations added for the CTI League included

- A full set of social media objects
- A one-line command to push information about a social media artifact up to MISP.

7.1.2 Related Work

CogSecCollab leads also helped start and chair the DEFCON AI Village (a village dedicated to work on the interface between information security, machine learning and artificial intelligence). One of the pieces of work aided by CogSecCollab was the 2019 Rootzbook misinformation challenge, designed as a simulation exercise to help young hackers understand the processes behind disinformation and botnets.

8 Sharing Disinformation Data with AMITT

8.1 Coordinating Responses

8.2 We need to tie this all together. Whole-system attacks often need whole-system responses. We've seen campaign creators use different types of accounts (bot, troll, cyborg, 'useful idiots' etc) across multiple platforms, topics and geographies; responses need to be across platforms, and will often be a mix of different blue team TTPs. This will need coordination across different groups, potentially through disinformation SOCs and ISAO-like bodies.

8.3 Making Tactics and Techniques Easy to Share

Online disinformation doesn't exist in a vacuum. The same types of framework that help campaign creators can also help with their removal. For instance, the easy access to demographic datasets that make micro targeting easy could be countered with stronger use of privacy laws and individual counters against online privacy invasions.

8.3.1 Sharing Formats

Checking parent models is also useful because this gives us formats for our counter objects—which is basically that these are of type "mitigation", and contain a title, id, brief description and list of techniques that they address. Looking at the STIX format for course-of-action gives us a similarly simple format for each counter against tactics — a name, description and list of things it mitigates against.

We want to be more descriptive whilst we find and refine our list of counters, so we can trace our decisions and where they came from. A more thorough list of features for a counter list would probably include:

- id
- name
- brief description
- list of tactics can be used on
- list of techniques can be used on
- expected action (detect, deny etc)
- who could take this action (this isn't in the infosec lists, but we have many actors on the defence side with different types of power, so this might need to be a thing)
- anticipated effects (both positive and negative — also not in the infosec lists)
- anticipated effort (not sure how to quantify this — people? money? hours? but part of the overarching issue is that attacks are much cheaper than defences, so defence cost needs to be taken into account)

And be generated from a cross-table of counters within incidents, which looks similar to the above, but also contains the who/where/when etc:

- id
- brief description
- list of tactics it was used on
- list of techniques it was used on
- action (detect, deny etc)
- who took this action
- effects seen (positive and negative)
- resources used
- incident id (if known)
- date (if known)
- counters-to-the-counter seen

9 AMITT for disinformation analysis

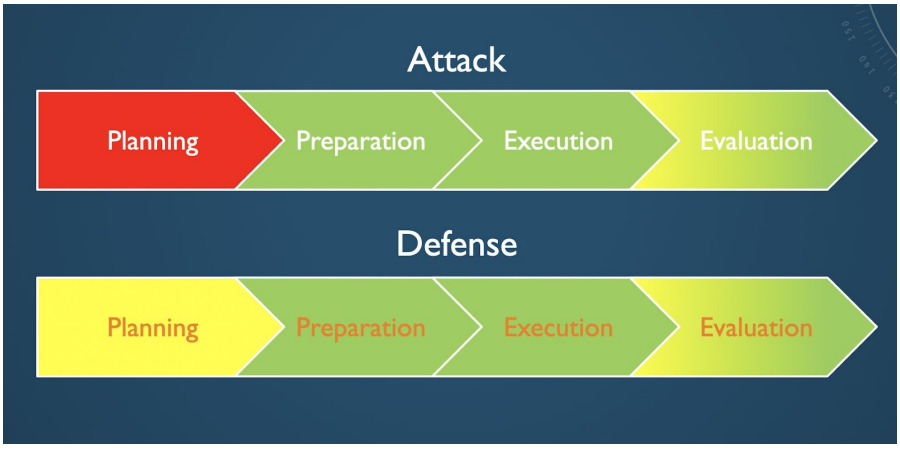


Figure 9.1 AMITT Phases and automation

9.1 Introduction

One of the advantages of the AMITT framework is that it allows an analyst to drill down to the most minute details of how a tactic, technique, or procedure is executed and can be countered while simultaneously allowing decision makers to make strategic decisions without becoming bogged down in the minutia of an individual action. In the early stages of planning, strategic thinkers may ask if a particular technology is the right fit for a given plan. Unfortunately, it's more likely that a decision maker will attempt to shoehorn a technology to their particular use-case.

This chapter will examine how a technology, in this case artificial intelligence (AI), can be examined for use in misinformation attacks and defense. The examination will be conducted from the vantage point of a strategy as opposed to a tactical action. In this context, strategy is an idea which guides employment of numerous resources or actions to achieve a desired end-state. A tactical action, by comparison, is the employment of resources and actions to support the achievement of a larger strategy. To summarise, tactical actions are time and resource limited and are in support of a larger, overarching strategy. One of the advantages of strategic thinking is it allows a planner to abstract away a lot of details and niche cases. At a minimum, this kind of strategic analysis helps to identify what things may be of most use and seed further detailed discussions.

At the strategic level, AMITT identifies four phases for a misinformation operation: planning, preparation, execution, and evaluation (Figure 9.1). The rest of this chapter will examine what each phase entails and how AI may be used to automate actions in that phase for both offense and defense.

9.2 Planning Phase

9.2.1 Offensive Misinformation

The planning phase requires two tasks to be completed: strategic planning and objective planning. Strategic planning is an inherently human decision and centers around the question of why the misinformation operation is being conducted. If the operation is successful, what is to be gained or accomplished? As strategic goals are a human desire for change, automation is not of use and AI is not well suited to assisting in this task.

Objective planning requires that a center of gravity analysis be conducted on the target. It is critical to note that there is one center of gravity. The center of gravity is the keystone from which every other facet of the target grows. Identifying the center of gravity for a target population requires in-depth understanding of society being targeted. Social norms, community norms, biases, and identity help form and inform the center of gravity. This kind of social analysis cannot be directly automated.

Where automation may help is in gathering large amounts of data to allow for broad scoping. By using automation and AI it's easy to collect information which will allow for identification of subjects of interest to the population as well as sentiment analysis for the population. This data will require further qualitative analysis to include "binning" before it is useful.

9.2.2 Countering Misinformation

One of the advantages defenders have in the battle of misinformation is their knowledge of themselves, their groups, and their society. Defenders know their own center of gravity and they know what must be protected. Likewise, defenders know what functions and objectives are critical and what normal behavior looks like making detection of anomalous activity feasible. As such, AI becomes a plausible tool in the planning phase by enabling the detection of anomalous activity and serving as an early warning system of possible attack.

9.3 Preparation Phase

9.3.1 Offensive Misinformation

The preparation phase is the phase where necessary resources are developed. Network development in the preparation phase includes understanding existing social networks as well as cultivating "useful idiots" to unknowingly propagate misinformation payloads. AI can automate these analyses through the use of social media analysis. AI is also useful for microtargeting. As a matter of fact, social media services are purpose-built for micro-targeting of advertisement; their entire business model is built on this capability. If we ignore laws about false advertising, the difference between advertising and misinformation is merely intent.

AI can also be tremendously useful in generating and honing message content. While the use of AI to generate deep fake videos is well-discussed in the open press, the use of systems like GPT-3 (Generative Pre-trained Transformer 3) to generate text narratives has gone largely undiscussed outside of academic circles. Further, using the advertising tools provided social media networks, AI has become remarkably adept at conducting A-B testing to determine which of two messages is most effective and using that information to generate the next iteration of messaging content.

9.3.2 Countering Misinformation

The defensive tactic most often discussed to counter misinformation is to provide counter narratives. In this sense, defenders can take advantage of automation in many of the same ways as attackers. Network analysis will allow for the identification of influencers which may help populate and transmit positive narratives. Network analysis also helps identify which groups may be susceptible to misinformation campaigns and therefore may be good targets for a prophylactic round of messaging by defenders. Finally, AI can be useful for A-B testing and honing of defensive messaging.

9.4 Execution Phase

9.4.1 Offensive Misinformation

The execution phase is where messaging reaches target audiences. AI can be used to identify existing and emerging influencers for pump priming. The use of AI bots to amplify misinformation is, however, a mistake as numerous computation methods exist for identifying bots and the messages they're currently spreading. Spotlighting attack infrastructure is a sure way to let defenders get the upper hand and deny needed resources to attackers. Smarter chatbots enabled by technologies such as GPT-3 can automate answering challenges to messaging as well as responding to queries from targeted audiences. Finally, AI is useful in continued A-B testing as well as in the effort to persist at the forefront of the target audience's mind.

9.4.2 Countering Misinformation

AI is of use in countering a misinformation campaign in its execution phase. Most current AI efforts concentrate on the use of AI to identify misinformation, but don't address the psychological facts of audiences having ingested misinformation. Misinformation works, in part, because it plays to pre-existing biases in the target audiences. By the time a target audience has been infected with misinformation, defenders must overcome cognitive dissonance, cognitive friction, and likely cognitive easing. This means that simply pointing out something isn't true is unlikely to have a significant effect.

The use of AI for fact checking has presented challenges, but does not make it useless. Considerations of whether the AI should assume an open-world or closed-world system will provide vastly different results and may be suitable in some cases. Likewise, known models for cascade-based and time-based propagation of misinformation are suitable for automation in AI systems. The use of AI to create deepfakes produces numerous anomalies that may be detectable by an AI. Mathematical anomalies in file content created by their creation are easily detectable by AIs. Likewise, medical anomalies can be detected by AIs. Micro-changes in complexion due to the heart pumping are easily detectable by AIs and the lack of such in a regular rhythm is a good indicator of altered media.

9.5 Evaluation Phase

9.5.1 Offensive Misinformation

As offensive campaigns continue, it is necessary to measure the effectiveness of ongoing efforts. AI is useful for gathering large amounts of data for measuring effectiveness and again for A-B testing of parallel efforts. While AI bots may be used for keeping the existence messages persistent, they are as easily detectable as they were in earlier phases. Worse, it's likely that now that the attack has executed, defenders have been alerted and are now actively looking for signs of attacker infrastructure. AI enabled evaluation enables for a rapid evaluation of current efforts and generation and testing of new content for rapid execution of additional iterations.

9.5.2 Countering Misinformation

The use of AI by defenders in the evaluation phase is useful in all of the same ways that it is for attackers. AI can enable the collection and evaluation of measures of effectiveness, help conduct A-B testing of counter narratives, and help speed the next iteration of the next effort. One advantage is that defenders don't necessarily mind if their infrastructure is highlighted to attackers as they have at least perceived legitimacy. There is a danger, however, that the discovery of bots to spread information may be leveraged by an attacker by re-contextualizing the narrative in saying that legitimate authorities are trying to muffle dissenting voices.

9.6 Conclusion

While the AMITT framework allows for highly detailed, component-wise analysis and countering of misinformation attacks, it also allows for the abstraction of details for a more holistic analysis of misinformation attacks and defense at the strategic level. This strategic level analysis may be used to seed discussion of the appropriateness of tools in conducting/countering misinformation as well as the efficacy of such tools at various phases of the misinformation kill-chain.

10 Using Tools with AMITT

AMITT is designed to be used with a suite of other disinformation tools and processes, including the MISP open-source threat intelligence toolset. This section covers disinformation adaptations to MISP and other toolsets.

10.1 MISP for Disinformation Tracking

MISP is used to store and share the artifacts and indicators of Cognitive Security incidents.

CogSecCollab have added the Misinformation Pattern Galaxy (AMITT), the DFRlab Dichotomy of Disinformation Pattern Galaxy, and other cogsec-related object templates - facebook-post, twitter-post, etc to MISP.

Objects in MISP can be thought of as STIX Domain Objects (SDOs). Relationships in MISP can be thought of as STIX relationship objects (SROs).

10.1.1 MISP Object Templates

The MISP objects we've created, and prior objects relevant to disinformation, are listed below. The code for MISP objects, the AMITT framework in MISP, and the DFRLab Dichotomy of Disinformation can be found in the MISP Project GitHub repo.

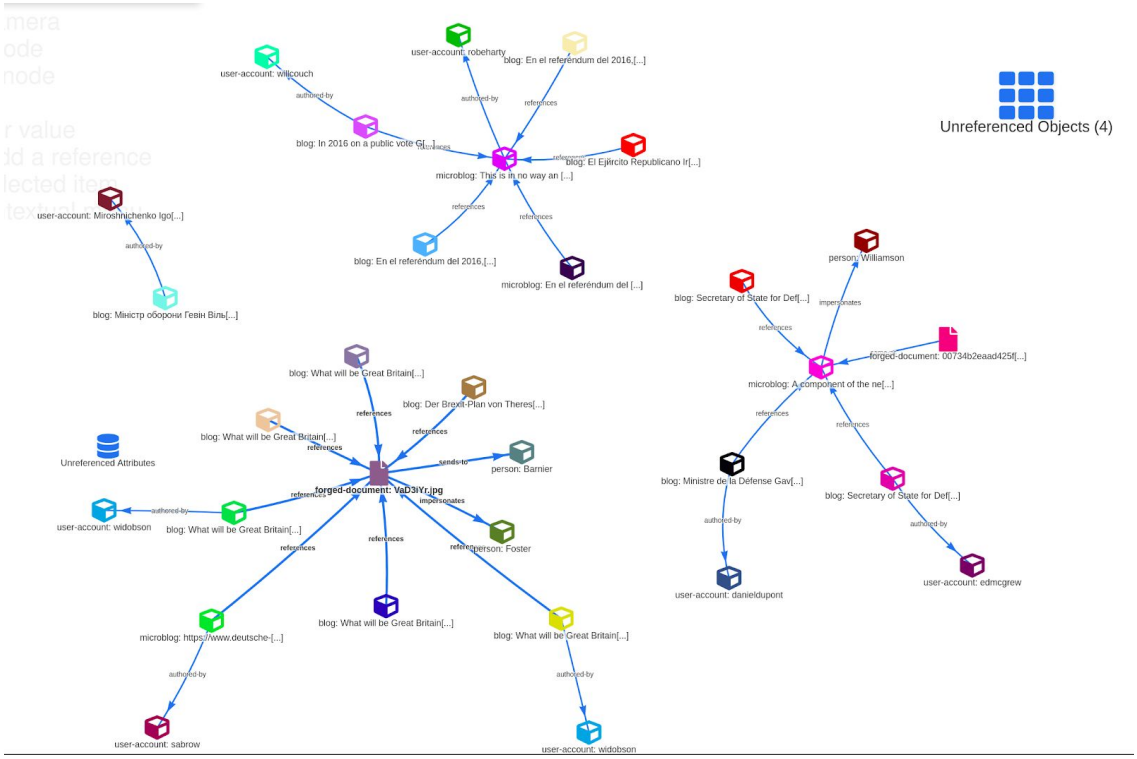
Object	**Misp**
Facebook group	misc:facebook-group
Facebook page	misc:facebook-page
Facebook account	misc:facebook-account
Facebook post	misc:facebook-post
Twitter account	misc:twitter-account
Twitter list	misc:twitter-list
Twitter post	misc:twitter-post
Blogsite	network:url
Blog account	misc:user-account
Blogpost	misc:blog
Reddit group (subreddit)	misc:reddit-subreddit
Reddit account	misc:reddit-account
Reddit post	misc:reddit-post
Reddit post comment	misc:reddit-comment
YouTube Channel	misc:youtube-channel
YouTube Video	misc:youtube-video
YouTube Playlist	misc:youtube-playlist
YouTube Comment	misc:youtube-comment
Website address	network:url
Instant message	misc:instant-message
Instant message group	misc:instant-message-group

Narrative	misc:narrative
Image	file:image
Meme	file:meme-image
Individual	misc:person
Event (e.g. protest)	misc:scheduled-event
Location	misc:geolocation

10.1.2 MISP Relationships

MISP Object relationships (think STIX Relationship Objects) define the relationship between all objects. These relationships let us describe how the pieces fit together. MISP relationships are found here: https://github.com/MISP/misp-objects/blob/main/relationships/definition.json

We can graph MISP relationships as shown below. This is one of the incidents in the Secondary Infektion campaign.



We can modify the objects/relationships to fit our specific needs. Since the MISP Project typically accepts PRs within a few days, it makes iterating over a data model fairly painless.

10.1.3 Example: The Narrative Object

The Narrative object stores a description of a narrative ("Bill Gates' C19 vaccines contain mind-control microchips") but it does not tell us anything about how that narrative relates to other entities (blog posts, persons, merch, etc).

So why use a Narrative object? It lets us store more information than MISP event tags, and keeps the UI cleaner (less wear on analysts). Artifacts within an incident might relate to different narratives, and we want to be deliberate in communicating which objects belong to what. What issues might you face with using narratives (and any other object) this way? Objects are a set of attributes and all attributes are automatically correlated via literal string matching. This works well for infosec artifacts but is more complicated with influence operations where totally unrelated campaigns might routinely reference the same persons, sites, etc. The immediate challenge with using Narratives is that in order to correlate two distinct incidents we must use a standard Narrative definition. This is an open problem.

10.2 STIX Viewers

STIG (https://github.com/idaholab/STIG) is a useful GUI-based tool for drawing and sharing STIX graphs outside the larger tools like MISP.
Stixview https://github.com/traut/jupyter-widget-stixview is a useful way to display STIX graphs from within Jupyter/ Python.

11 Further Work

Uses include the use of NLP, social graph analysis, propagation patterns on raw data. Lots of approaches that are only pieces of the puzzle, or intractable/unscaleable. Difficulty of counteracting entrenched beliefs directly, information aikido, disrupting the coordination of meme/conspiracy attacks. Importance of information sharing for detecting campaigns early. AMITT, kill chains, counterterrorism models. Potential for AI/ML approaches to detection and automated countermeasures.

Write more about the DE of ABCDE and how it links to AMITT risk management models.

Add Anti-harassment models to counters section.

Convene users and designers to work through the proposed changes to AMITT.