Analyzing frameworks to model disinformation attacks in online social networks

Gonzalo Cánovas López de Molina*, Felipe Sánchez González*,
Pantaleone Nespoli*, Javier Pastor-Galindo*, José A. Ruipérez-Valiente*
*Department of Information and Communications Engineering, University of Murcia, 30100, Murcia, Spain
{gonzalo.canovasl, felipe.sanchezg, pantaleone.nespoli, javierpg, jruiperez}@um.es

***Abstract*—Disinformation campaigns are increasingly prevalent tactics employed by various actors to advance their agendas, underscoring the critical imperative to bolster detection systems. Consequently, the establishment of a comprehensive system becomes paramount, facilitating the categorization of these attacks and the discernment of perpetrators' methods in disseminating disinformation, thereby exerting influence over public opinion. This paper embarks on an extensive examination of prominent disinformation frameworks, including ABCDE, BEND, ALERT, SCOTCH, and DISARM. Furthermore, it conducts a detailed analysis of a real-world case study centred on the Mali scenario, leveraging three frameworks to glean invaluable insights into malicious tactics and enhance classification capabilities. Additionally, the paper advocates for the adoption of an additional framework to formulate robust countermeasures against future disinformation campaigns. The significance of these findings cannot be overstated, as they are instrumental in comprehending and categorizing such attacks, thereby enabling proactive measures to forestall similar occurrences or the exploitation of communal resources by nefarious entities advancing their agendas.**

***Index Terms*—Misinformation, Disinformation, Online Social Networks, Cybersecurity, Cyberdefence**

Type of contribution: *Original research*

I. Introduction

The rise of the use of social networks in the last decade has produced a huge impact on today's society. Nowadays, people from all around the globe use their devices to inform themselves on platforms such as TikTok, X (former Twitter), or Instagram. On one hand, these platforms empower organizations and anonymous actors to inform society in a dynamic way but, on the other hand, it enables malign actors to distribute their propaganda, potentially allowing a shaping of society's opinions, as has been seen recently with the Ukraine war [1].

These malign actions are known to politicians and rulers, who make joint efforts to combat disinformation threats and embrace the use of structured and more formal ways to detect them, allowing for better and quicker responses [2]. These matters are especially relevant considering that the world approaches a multiple-election year, which could potentially be a target for disinformation threats [3], [4].

Moreover, disinformation frameworks have been proven to be capable tools to combat disinformation campaigns, not only for mature practitioners who have a settled knowledge but with other more novice users who want to adopt them in their workflow by providing easy and intuitive associations [5].

However, upon investigating the current literature covering this topic, there is a lack of extensive review of frameworks for disinformation modelling and, especially, practical application in real use cases.

In particular, the objectives of this study are the following:

1) Identify and compare existing frameworks to model disinformation attacks, which can offer responders various alternatives for addressing the inherent challenge of detecting and classifying ambiguous disinformation threats
2) Model a real-world disinformation attack in a cyberdefence context using different analyzed frameworks. This approach will provide a practical demonstration of the frameworks' performance in a real-world scenario.

With these objectives in mind, firstly, Section II provides an overview of the current state of the art in disinformation modelling. After, in Section III, a review of the most interesting frameworks found to model disinformation is provided, comparing their main features, their advantages and disadvantages. Then, Section IV describes a real-world example where a disinformation campaign is presented. Finally, the reviewed frameworks are applied to the presented campaign see their real performance when addressing real disinformation threats.

II. State of the Art

Understanding and combating disinformation today poses a significant challenge [6]. The widespread dissemination of false information across digital platforms has made it increasingly difficult to discern truth from falsehood, particularly as it spreads rapidly through social networks and mobile apps.

Recent research has turned to mathematical modeling to develop effective countermeasures [7]. By analyzing real social network data, scholars have explored the impact of content moderation, education, and counter-campaigns. Surprisingly, indiscriminate removal of disinformation sources, regardless of their influence, can rival targeted approaches. Strategies fostering public scepticism and engaging widespread participation in counter-campaigns have shown promise.

The Online Misinformation Engagement Framework offers a structured approach to understanding the complexity of misinformation engagement [8]. By delineating key stages such as source selection and information evaluation, interventions targeting these early stages are seen as crucial yet underexplored.

In response to evolving tactics, continuous monitoring and detection mechanisms are essential [9]. Workflows designed to detect coordinated social media actors have proven vital in identifying emerging threats, particularly during critical events like elections.

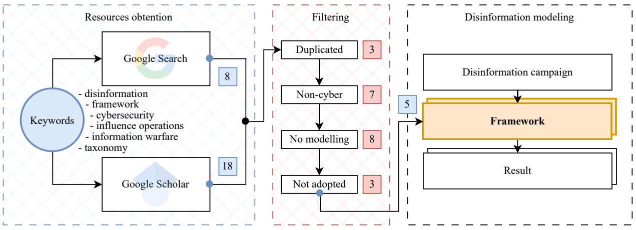

Fig. 1. Followed methodology to obtain the proposed frameworks

Additionally, addressing disinformation requires a comprehensive understanding across platforms and products [10]. The ”ABC” framework highlights the intertwined nature of manipulative actors, deceptive behaviors, and harmful content, calling for interdisciplinary and collaborative strategies to combat misinformation effectively.

Other works have proposed to model disinformation attacks using already existing and widely adopted cybersecurity models, thus bridging those ecosystems.

In [11], a cybersecurity-inspired framework was proposed to model disinformation attacks and mitigations. The authors engaged 22 experts on disinformation through interviews and qualitative coding, identifying domains, functions, and threat features for the disinformation landscape. Drawing from the cybersecurity ecosystem, the approach demonstrated viability when applied to real-world disinformation campaigns, highlighting the need for automation due to the predominantly human resources dedicated to disinformation protection.

Similarly, [12] introduced a Cyber Kill Chain (CKC) approach to understanding influence operations in social media, breaking down the process into seven phases. By combining socio-technical perspectives, the study shed light on often overlooked elements, emphasizing the importance of social media's characteristics in facilitating influence operations.

In [13], a high-level disinformation model was presented, extending the concept of Advanced Persistent Threat (APT) to include strategic, operational, and socio-organizational levels. The Advanced Persistent Threat Operational Line (APTOL) framework and a disinformation model based on Situational Awareness (SA) theory were proposed to counteract adversaries passively.

Lastly, a NATO report [14] addressed the need for an officers' training platform on Influence Operations (IO), outlining a global IO model for simulation. The report defined operation concepts and presented an IO training platform, serving as a proof-of-concept for future full-fledged IO simulation platforms.

These contributions highlight ongoing efforts in academia, the military, and the industry to develop effective models and methodologies to defend against disinformation attacks, with frameworks like AMITT (Adversarial Misinformation and Influence Tactics and Techniques) [15] and DISARM [16] gaining traction in the field.

III. Frameworks for disinformation modelling

To survey the existing frameworks modelling disinformation in the literature, the process shown in Figure 1 was followed, as described next.

To initiate the search process, specific keywords were carefully chosen. These included “disinformation,” “framework”, “cybersecurity”, “influence operations”, “information warfare”, and “taxonomy”. Subsequently, the search was conducted utilizing Google Search and Google Scholar, producing 8 and 18 resources respectively. Thirdly, the frameworks had to comply with the following characteristics: to not be duplicated in the search, to follow a cyberdefense or cybersecurity approach, to model disinformation, and to be embraced by some important organizations. Finally, after analyzing the resources and applying the filters, 5 frameworks were selected to analyze the proposed disinformation campaign.

A. *DISARM framework*

The DISARM framework [16] is a project maintained by the DISARM foundation that aims to offer a standardized way of modelling disinformation incidents. DISARM uses the Cyber Kill Chain and Tactics, Techniques and Procedures (TTP) to categorize the different stages of a disinformation incident, giving the framework a cyber defensive facet. Apart from these stages, it is designed to be interoperable with the ATT&CK framework [17] which is widely used in the information and cybersecurity community. This interoperability is achieved by using its own STIX/TAXII-based data format, adopting a cybersecurity approach and making disinformation incident information interoperable between the already available STIX and TAXII solutions.

Specifically, it is composed of two frameworks:

1) The **DISARM Red framework**, which presents a repository of TTP that potential disinformation threat actors could utilize to propagate disinformation narratives against a collective or state.
2) The **DISARM Blue Framework**, which proposes a series of countermeasures to mitigate disinformation incidents. This framework provides a direct mapping between its items and the Red framework ones, allowing for an intuitive and agile incident lessening.

Additionally, these frameworks are composed by three main components:

1) **Phases**, which are the most abstract groups of tactics and their techniques that take place in a certain stage of the influence campaign, i.e., Plan, Prepare, Execute, and Assess.
2) **Tactics**, which are how the objectives of the campaigns are achieved. For example, for the Prepare phase, the following tactics are found: Develop Narratives, Develop Content, Establish Legitimacy, etc.
3) **Techniques** (countermeasures in the Blue framework), which tell how a particular tactic has materialised. They can be associated with one or more tactics as one technique may serve to achieve different goals. Some techniques also contain subtechniques that provide more concretion about how the tactic was performed.

To understand the interconnection of these elements, consider the example of the “Develop Memes” subtechnique. This subtechnique falls under the broader “Develop Image-Based” technique, which contributes to achieving the “Develop Content” tactic within the Prepare phase.

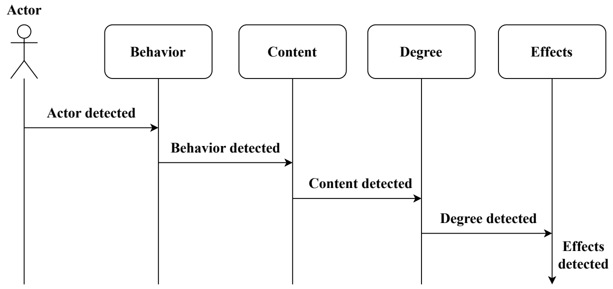

Fig. 2. Diagram of steps to use ABCDE framework

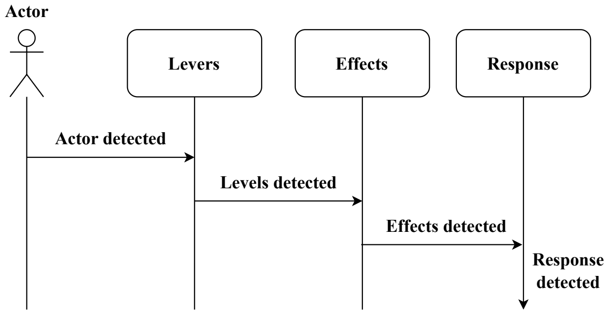

Fig. 3. Diagram of steps to use ALERT framework

Furthermore, DISARM provides certain real-life incidents that were associated with some of the techniques that were used in it. With this association, responders and creators are provided with examples of how some techniques were implemented and the resulting consequences obtained.

The framework offers significant advantages including an information security approach to disinformation, a comprehensive repository of disinformation techniques, direct mapping of attacking techniques with their countermeasures, and integration of STIX/TAXII data formats for compatibility with existing solutions. However, it is accompanied by certain drawbacks, such as a highly technical perspective on disinformation, incomplete countermeasure mapping for some attack techniques, and a lack of guidance on evaluating the impact of techniques and countermeasures.

B. *ABCDE framework*

The ABCDE framework [18] was created in 2020 to address the growing threat of online disinformation. It was developed by James Pamment, at the Carnegie Endowment for International Peace.

The ABCDE framework stands as a robust analytical tool for dissecting the multifaceted challenges of disinformation and foreign interference. Its five key components are actor, behaviour, content, degree and effect.
Firstly, by scrutinizing the diverse entities involved, from individual actors to foreign states, it lays the groundwork for understanding motives and potential sources of influence.
Secondly, it delves into the tactics employed by these actors, shedding light on the transparency, authenticity, and intent behind their actions.
Moving on to the content aspect, it evaluates the narratives and language used, providing insights into the harmfulness and potential impact of disseminated information.
Then, measuring the scale and reach of these activities, aids in prioritizing responses and resource allocation.
Finally, by assessing the societal impact, including effects on public discourse, trust, and national security, it guides policymakers in developing targeted interventions to counteract these threats effectively.

The process shown in Figure 2 was followed to have all the classifications.

C. *ALERT framework*

ALERT [19] was proposed by QUT Business School, University of Melbourne's School of Computing and Information Systems and finally Institute for Defence Studies and Analyses, New Delhi.

This framework stands as a significant academic contribution to comprehending and tackling political manipulation through information systems. Its robustness lies in its holistic approach, offering a nuanced understanding of this intricate phenomenon and guiding research and intervention effectively.

Primarily, it meticulously delineates the diverse actors engaged in politically motivated information manipulation, acknowledging their pivotal role in shaping societal and political narratives. Additionally, it adeptly dissects the mechanisms and strategies employed by these actors (Actors) to influence public perception and political outcomes, from misinformation dissemination to algorithmic manipulation (Levers).

Furthermore, the framework rigorously examines the societal and political ramifications of information manipulation, encompassing social polarization, erosion of trust in democratic institutions, and disruptions in electoral processes (Effects). Lastly, it furnishes practical guidance for formulating responsive measures (Response Taxonomy), aiding in the formulation and implementation of effective policies to mitigate adverse effects and uphold the integrity of democratic processes.

The process shown in Figure 3 was followed to have all the classifications.

D. *BEND framework*

BEND [20] is a framework developed by a team of researchers at Carnegie Mellon University as part of the Social and Semantic Systems (S3D) project, also having the US Army as a contributor.

The BEND framework is a powerful analytical tool for studying and combating misinformation online. Its comprehensive structure meticulously examines the various aspects of online misinformation, offering deep theoretical insights. It serves as a guiding light for research, leading to a more complete understanding of how misinformation spreads and is received in the digital realm.

This framework outlines two key dimensions: manipulating narrative and social networks.

In narrative manipulation, objectives range from positive actions like engaging audiences and enhancing discussions to negative tactics such as dismissing issues or distorting messages. These aims shape discourse to align with the intentions of those spreading misinformation. Likewise, in social

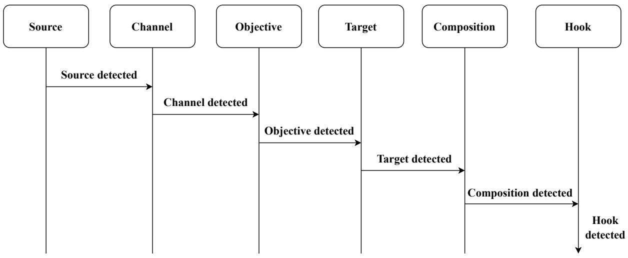

Fig. 4. Diagram of steps to use SCOTCH framework

network manipulation, objectives aim to strengthen or weaken connectivity and influence. Positive goals include boosting opinion leaders and fostering group cohesion, while negative goals involve neutralizing leaders or isolating factions. These tactics control how misinformation spreads and is received, utilizing online social structures to amplify or diminish its impact.

This framework uses the four E's for positive objectives (engage, explain, excite and enhance) and the four D's for negative objectives (dismiss, distort, dismay and distract). This is only for the manipulation of narratives. On the other hand, for manipulation of social networks, the framework uses the four B's for positive objectives (back, build, bridge and boost) and for negative objectives, the four N's (neutralize, nuke, narrow and neglect).

E. SCOTCH framework

SCOTCH [21] is a framework for rapidly assessing influence operations. It was developed by researchers at the Atlantic Council.

This framework renowned for its discerning ability in the realm of influence operations, relies on a series of essential elements that delve deep into understanding such phenomena. Among these elements, the precise identification of the source of information (Source) stands out, serving as the point of origin for the manipulative narrative. Through the analysis of the communication channels (Channel) used to disseminate the message, strategies of distribution and reach of distorted information are revealed. The overarching objectives (Objectives) and specific goals (Goals) delineate the aims pursued by the architects of the operation, providing a clear insight into their intentions and motivations.

On the other hand, the composition of content (Composition) emerges as a critical component, outlining the format, style, and message used to influence the audience. This aspect is complemented by the consideration of "hooks", persuasive techniques designed to capture and retain the recipient's attention, constituting a fundamental pillar in the effectiveness of the influence operation. Together, these elements form a solid and coherent structure that allows for the breakdown and comprehension of the complex dynamics of influence operations in the digital realm.

The process shown in Figure 4 was followed to have all the classifications.

F. Summary

In short, the analysis of these frameworks provides large amounts of information, such as: who the actors are, who they attack, and why, among other things.

The frameworks chosen for the classification of attacks were ALERT and ABCDE due to their ease of finding information, the help with questions to detect each dimension (e.g. actors, behaviour, among others) and their examples of classification of other attacks, which also help to understand how it is possible to use them. On the other hand, the SCOTCH and BEND frameworks have been discarded for classification due to their complexity of usefulness, the fact that the classification does not give enough information to detect certain dimensions and sometimes it can be confusing to detect them, compared to the previous ones where they present as "a guide" to detect each one. In turn, DISARM has been chosen as a framework to see the types of attacks carried out and the countermeasures to be applied for each attack.

Each framework provides information about the actors, in some cases in a more detailed way such as the ALERT and ABCDE frameworks, and all of them offer example use cases to see how a real case would be analysed, which is interesting to see how the frameworks work when using them. They also offer a classification of disinformation attacks (to classify the frameworks) and for the quantitative analysis part (offering values for classification) only the BEND framework offers such an analysis, which can be used to evaluate each part. On the other hand, the DISARM and BEND frameworks offer a series of stages to detect the phases of an attack, which is important to understand the attack; the classification frameworks offer dimensions that detect certain aspects such as actors, and behaviour, among others. SCOTCH and BEND offer certain dimensions that are complicated to classify when there is a disinformation attack; on the other hand, the ALERT and ABCDE frameworks offer more clarifying dimensions when classifying such an attack. On the cyber side, DISARM relies on Mitre Attack and CyberKill Chain to work. Finally, they are all used by public organisations, while the ALERT framework is used for research.

In Table I, a summary of the discussion is presented

IV. Modelling a real-world disinformation attack: The case of Wagner at Mali

This Section describes how the selected frameworks would model a real use case to learn how each one brings relevant information to understand the attack.

A. Use case description

In late 2021, Russian Wagner mercenaries arrived in Mali to allegedly combat terrorism in the region. Preceding the arrival, an important amount of Facebook pages promoting pro-Russian and pro-Wagner propaganda increased their presence on the platform [22], [23]. In other words, a disinformation attack was designed and launched to support and complement the kinetic activities.

A network of Facebook pages operating from Mali emerged as a conduit for disseminating pro-Russian and anti-French narratives, particularly in the context of the presence of Wagner Group mercenaries in Mali and the aftermath of a coup in May 2021. These pages, posing as charitable and community-oriented, strategically aimed to undermine French interests, advocate for Russia as an alternative to Western

Features	DISARM	SCOTCH	BEND	ABCDE	ALERT
Proposed by	DISARM Foundation	Atlantic Council	Carnegie Mellon and US Army	Carnegie Endowment for International Peace	QUS Business School, University of Melbourne and IDSA
Actors analysis	-	✔	✔	✔	✔
Use case examples	✔	✔	Chapter 5 of its paper	UE examples	✔
Supported by	UE, OTAN, ONU	-	-	UE	-
Disinformation incident classification	✔	✔	✔	✔	✔
Countermeasures provision	✔	-	✔	✔	✔
Quantitative analysis	-	-	✔	-	-
Stages	Plan, Prepare, Execute, Assess	-	Framework workflow	-	-
Dimensions provided	-	Source, Channel, Objective, Target, Composition, Hook	[Narrative / Social Network] & [Negative / Positive] maneuvers	Actor, Behaviour, Content, Degree, Effect	Actor, Levers (with levels), Effects, Response
Cyber analogy	MITRE ATT&CK and CyberKill Chain	-	-	-	-
Research or Public Organisations	Public Organisations	Public Organisations	Public Organisations	Public Organisations	Research

TABLE I
SUMMARY OF FRAMEWORKS ANALYSED

influence, and garner support for interim President Assimi Goïta and the Malian military [23].

The coordinated efforts of these Facebook pages involved posting content across multiple platforms simultaneously and sharing contact information that linked various events, suggesting a cohesive network behind them. In November 2021, a new page focused on the broader Sahel region was introduced, advocating for a "revolution" while amplifying the pro-Russian and anti-French rhetoric seen across the network [23].

This incident was not isolated, as it mirrors previous instances of social media manipulation in Africa. In December 2020, Facebook dismantled a French network targeting Mali with anti-Russian messaging, along with competing French and Russian networks engaging in similar activities in the Central African Republic. These actions highlight the proxy war between French and Russian interests in the region, with social media serving as a battleground for shaping narratives and influencing public opinion [23].

B. Attack modelling

1) ABCDE Framework: The application of the ABCDE frameworks enables the identification of the actors, behaviour, content, degree and effects of the Wagner disinformation attack in the context of the Mali conflict.

First of all, it is needed to detect who is the actor in this attack. The ABCDE framework gives some questions to ask and find the Actor like "*Is the actor affiliated with a private or nongovernmental organization?*" or "*Is the actor an agent or proxy of a foreign government?*" According to documented literature, it is highly probable that the responsible for this disinformation attack is a governmental organization of Russia.

Secondly, it is needed to detect the behaviour of the attack, and as above, the framework gives some questions to ask like "*Is the actor disguising his or her identity or actions?*" or "*Is there evidence of back-end coordination?*". According to the evidence, the attack was hiding information and giving false information to disseminate pro-Russian and anti-French narratives and promoting pro-Russian and pro-Wagner propaganda on Facebook.

After that, the framework analyzes the content used to give some information such as the nature or severity.As mentioned above, it is needed to answer some questions that will help to find the real content like "*Which languages are used in the spread of the disinformation or other online content in question?*" or "*Is the content harmful?*" After analyzing the evidence, it becomes apparent that the content in this particular use case involves disseminating false information to manipulate public opinion to garner Western support for interim President Assimi Goïta and the Malian military.

The framework characterizes the degree of the attack which is going to give some information about examining information regarding the dissemination of the content in question and the audiences it reaches. As previously indicated, it is necessary to address certain questions such as "*Is the content going viral on social media platforms in a way that would suggest an inauthentic boost to online engagement?*" or "*Is the content tailored or microtargeted, and, if so, to which audiences?*" After addressing these inquiries, it is apparent that the Degree section involves orchestrating a campaign to sway the people of Mali and propagate a pro-Russian, anti-French ideology.

Finally, the effect is scrutinized to ascertain the extent to which a particular case presents a threat. Analogous to the aforementioned, it is imperative to pose certain inquiries that will facilitate the determination of said effects.

Several of these questions resemble "*Does the content dissuade voters from participating in elections or seek to undermine the results of an election?*" or "*Is the online content issue-based?*" Considering the questions posed earlier, it is evident that the degree of involvement in this particular use case encompasses undermining public trust, influencing public opinion, and destabilizing democratic processes in Mali.

After analyzing this use case with the ABCDE framework, it becomes apparent why the actor targeted their objective and what the actor sought to obtain.

2) ALERT Framework: :

The application of the ALERT framework enables the identification of the actors, levers (with four levels), effects and responses of the Wagner disinformation attack in the context of the Mali conflict.

First of all, the framework needs to find the actor in this use case. Once the actor is located, this actor can: has to be classified as one or more of these: instigate the attack, partially support any aspect of planning the attack, conduct the actual attack, support the concealment of the source or amplify the attack.

In this particular use case, the identified actor is a governmental organization from Russia, encompassing all conceivable categories.

Secondly, the Lever section gives information about how actors can leverage information systems to interfere and compromise.

There are four levels in this lever:

1) Disrupting sources of data in physical and logical systems: Involves intentionally disrupting data generation, collection, storage, or transmission in both physical and digital environments.
2) Manipulating algorithms used in the processing of signals: Involves intentionally altering algorithms to influence the interpretation, analysis, or transmission of signals within various technological contexts.
3) Manipulating interpretations associated with information: Altering the understanding or context of information. This can involve placing valid information in a false context or presenting false information within a true context.
4) Weaponizing information systems: Strategic deployment of digital platforms and communication channels with the intent to achieve political, physical, economic, or social impact, often resulting in direct harm or disruption.

In the hierarchical delineation of levels, the attribution of the Russian attack corresponds to level three. This determination stems from Russia's strategic manipulation of information to propagate a pro-Russian and anti-French ideological narrative.

After that, it needs to find the Effects caused by the attack, where the framework provides information on what effect the actors want to cause at the time they carry out the attack.

There are three kinds of effects: IW influence is characterized by the strategic shaping of perceptions and behaviours through tactics like propaganda and communication campaigns; IW interference, involves the deliberate disruption or manipulation of information systems and processes to undermine stability and integrity and IW hacking, which denotes unauthorized access and exploitation of digital systems for various objectives, such as espionage or sabotage.

Once these three typologies are delineated, the effects observed within the context of the disinformation attack on Mali manifest distinct manifestations:

Through IW influence, the propagation of pro-Russian and anti-French narratives via Facebook pages within Mali was directed towards the shaping of political sentiment and the cultivation of distrust towards the government. Conversely, within the realm of IW interference, particular instances, notably coinciding with events such as the deployment of Russian mercenaries to Mali, witnessed orchestrated information manipulation aimed at bolstering kinetic activities. This phenomenon suggests a concerted endeavour across multiple platforms, indicative of a coordinated strategic approach.

Finally, it needs to know the Response that brings information on the options accessible to nations targeted by IW. They can be categorized into five measures: IW defense focused on safeguard information systems and networks from unauthorized access and cyber threats; IW offense involves retaliatory actions against weaponized information systems, allowing for preemptive or retaliatory use of information warfare means; Diplomacy entails collaborative international endeavours to establish norms, regulations and agreements governing online activities and behaviour; Legal sanctions are penalties enforced by a state on sub-state entities to ensure adherence to the law and Economic sanctions involve commercial and financial penalties applied by one or more countries against the perpetrating party, aiming to deter or punish undesirable behaviour.

In this particular use case, IW-defense responses were exclusively employed to counter the attack. These responses encompassed strengthening verification methods, blocking sources of pollution, implementing a honeypot social community, identifying and removing or rate-limiting identical content, exposing actors and their intentions, creating friction by marking content with ridicule or other "decelerates," disavowing disinformation by respected figures, using humorous counter-narratives, removing or rate-limiting botnets, prebunking, and implementing social media amber alerts.

After analyzing this use case with the ALERT framework, the framework elucidates the rationale behind the actor's pursuit of their objective, delineating their motivations, desired outcomes, and the methods employed to achieve their ends.

3) DISARM Framework: : To show the DISARM Framework application, the obtained disinformation artefacts were associated with the different techniques proposed by DISARM. These techniques were selected by analyzing the different disinformation artefacts left by the campaign and by choosing the techniques that seemed more appropriate for each of them. As a result, Table II represents the final DISARM techniques detected in Mali's disinformation campaign. The table is divided into the different DISARM phases and, for each one, the different tactics and techniques associated with the incident.

In the Plan phase, where the vision of the attack is starting to be comunicated by the incident organizators, just one tactic stands out: TA02: Plan Objectives. This tactic is focused on setting certain objectives and planning the desired effects. It was achieved mainly with the T0002: Facilitate State Propaganda technique, materialized thanks to the organization of pro-Russian Facebook groups which were coordinated to leverage Russian sympathy messages to Mali's population.

Subsequently, in the Prepare phase, tactics that are executed to improve the conditions for the later actions are found. In this phase, the main associated tactic is TA06: Develop Content. To reach it, the main techniques are: T0085: Develop Text-based Content and T0086: Develop Image-based

Tactic	Technique	Reason
	PLAN	
TA02: Plan Objectives	T0002: Facilitate State Propaganda	Use of pro-Russian post in the organisations' Facebook pages
	T0075.001: Discredit Credible Sources	Boycotts towards French media
	T0066: Degrade Adversary	France hate posts
	T0136: Cultivate Support	First GPM posts advocating for Russian cooperation.
	PREPARE	
TA14: Develop Narratives	T0003: Leverage Existing Narratives	Leveraging the cabinet reshuffle topic
	T0022: Leverage Conspiracy Theory Narratives	Theories about France influence in the cabinet reshuffle
TA06: Develop Content	T0015: Create hashtags and search artefacts	Created hashtags: #vive_l_armee_malienne, #nous_sommes_wagner
	T0019: Generate information pollution	Fake information about France's influence shortly after the cabinet reshuffle
	T0085: Develop Text-based Content	Opinions about news and Wagner cooperation promotion
	T0086: Develop Image-based Content	Images promoting Wagner support
	T0087: Develop Video-based Content	Publishing of organisation's pro-Russian conferences
TA15: Establish Social Assets	T0007: Create Inauthentic Social Media Pages and Groups	Creation of groups to promote Russian/Wagner cooperation in Mali
	T0065: Prepare Physical Broadcast Capabilities	Use of ”information” and merchandising phone numbers
	T0092: Build Network	Different groups to promote the Wagner narrative
	T0096: Leverage Content Farms	Suspicious followers and reactions numbers
TA07: Select Channels and Affordances	T0104: Social Networks	Facebook as the main driver to promote disinformation
	EXECUTE	
TA09: Deliver Content	T0115: Post Content	Facebook posts across different groups.
TA17: Maximise Exposure	T0049: Flooding the Information Space	Groups posted content within 60 seconds from each post.
	T0118: Amplify Existing Narrative	Promotion of anti-French and pro-Wagner content.
	T0119: Cross-Posting	Same posts across different pages.
TA18: Drive Online Harms	T0048.001: Boycott/”Cancel” Opponents	Boycotts towards French media
TA10: Drive Offline Activity	T0017: Conduct fundraising	Suspicious deals with Russia
	T0057: Organise Events	Pro-Russian event organisation.
	T0061: Sell Merchandise	Merchandising sales by promoting a phone hotline.
	T0126: Encourage Attendance at Events	Promotion of Russian/Wagner support marches.

TABLE II
DISARM RED (CREATOR) TECHNIQUES DETECTED IN THE INCIDENT

Content. They include the fabrication of misleading text, narratives alignements, support collages and memes, among others. Complementary to these techniques, the T0015: Create hashtags and search artifacts and T0019: Generate information pollution are used to leverage the future post.

Finally, the Execute phase was reached by the pro-Russian groups. In this phase, the TA09: Deliver Content is used as the main driver to carry disinformation to Mali's population. By being feeded by the T0115: Post Content technique, it achieves to distribute all the mentioned content in the form of Facebook posts across various groups.

In opposition to the DISARM Red framework, once Table II was built, a shift was made towards the defensive approach. The focus was placed on compiling the countermeasures that could be applied to mitigate an attack with these characteristics. Therefore, the techniques obtained from the Red framework were aborded and a set of appropriate countermeasures that could address them were selected. As a result, Table III shows possible DISARM Blue countermeasures that could be applied to combat the corresponding detected Red techniques from Table II.

Tactic	Countermeasure	Mitigates
	PLAN	
TA01: Plan Strategy	C00012: Platform regulation	[T0002] [T0007] [T0022]
	C00016: Censorship	[T0002] [T0007] [T0022] [T0049]
	C00006: Charge for social media	[T0007] [T0015]
TA02: Plan Objectives	C00030: Develop a compelling counter narrative (truth based)	[T0002] [T0003] [T0022]
	PREPARE	
TA05: Microtarget	C00066: Co-opt a hashtag and drown it out (hijack it back)	[T0015]
TA06: Develop Content	C00071: Block source of pollution	[T0019]
	C00091: Honeypot social community	[T0049]
	C00074: Identify and delete or rate limit identical content	[T0019] [T0022] [T0049] [T0061]
TA07: Select Channels and Affordances	C00099: Strengthen verification methods	[T0007]
	EXECUTE	
TA08: Conduct Pump Priming	C00119: Engage payload and debunk	[T0022]
	C00115: Expose actor and intentions	[T0048]
TA09: Deliver Content	C00128: Create friction by marking content with ridicule or other ”decelerants”	[T0049] [T0057] [T0061]
	C00200: Respected figure (influencer) disavows misinfo	[T0022]
	C00211: Use humorous counter-narratives	[T0022] [T0057]
	C00123: Remove or rate limit botnets	[T0049]
	C00125: Prebunking	[T0002] [T0003] [T0022]
	C00126: Social media amber alert	[T0002] [T0003] [T0022] [T0057]

TABLE III
DISARM BLUE (RESPONDER) COUNTERS TO MITIGATE THE DETECTED RED TECHNIQUES

Beginning with the Plan phase, the C00012: Platform regulation can serve as a countermeasure to the T0002: Facilitate State Propaganda, T0007: Create Inauthentic Social Media Pages and Groups and T0022: Leverage Conspiracy Theory Narratives techniques. A basic example for the Mali's campaign could be regulating Facebook so no political content is exhibited.

Continuing with the Prepare phase, the C00074: Identify and delete or rate limit identical content countermeasure could mitigate the T0049: Flooding the Information Space technique. This could have been implemented by detecting similar posts shared within a certain threshold and preventing the user from posting them until some time.

Finally, in the Execute phase, the C00126: Social media amber alert countermeasure could combat several techniques through alerting the user about possible disinformation.

V. Conclusions and future work

In conclusion, this paper has illustrated how disinformation frameworks enable responders with a structured vision of disinformation campaigns, which not only enables decision-makers to be more conscious with their actions but ensures a better way of combating disinformation by improving incident response agility and effectiveness.

After discussing major disinformation frameworks, this paper introduces the disinformation camping by Russia at Mali to assess the applicability of these frameworks in real-world scenarios. Thanks to the application of each of the frameworks, consequent categorizations are shown such as high-level management of the incident, effects tagging, techniques associations, and, most importantly, countermeasures to mitigate the disinformation effects. These categorizations can equip disinformation responders, both at high and low levels, with reliable tools to manage and combat disinformation in current and future incidents.

As future work, the creation of a unified hybrid framework that could improve the performance of the studied frameworks remain pending, as well as the study of other use cases and frameworks that could provide other perspectives in the modelling of disinformation.

Acknowledgment

This work has been partially funded by the strategic project CDL-TALENTUM from the Spanish National Institute of Cybersecurity (INCIBE) and by the Recovery, Transformation and Resilience Plan, Next Generation EU.

References

[1] B. van Niekerk, "The evolution of information warfare in ukraine: 2014 to 2022." *Journal of Information Warfare*, vol. 22, 2023.

[2] European External Action Service's (EEAS) Stratcom, "1st EEAS report on foreign information manipulation and interference threats," 2 2023. [Online]. Available: https://www.eeas.europa.eu/eeas/1st-eeas-report-foreign-information-manipulation-and-interference-threats_en

[3] ——, "2nd EEAS report on foreign information manipulation and interference threats," 1 2024. [Online]. Available: https://www.eeas.europa.eu/eeas/2nd-eeas-report-foreign-information-manipulation-and-interference-threats_en

[4] U.S. Office of the Director of National Intelligence, "Annual threat assessment of the u.s. intelligence community," 3 2024. [Online]. Available: https://www.dni.gov/index.php/newsroom/reports-publications/reports-publications-2024/3787-2024-annual-threat-assessment-of-the-u-s-intelligence-community

[5] H. Newman, "Foreign information manipulation and interference defence standards: Test for rapid adoption of the common language and framework 'DISARM'," 11 2022. [Online]. Available: https://stratcomcoe.org/publications/foreign-information-manipulation-and-interference-defence-standards-test-for-rapid-adoption-of-the-common-language-and-framework-disarm-prepared-in-cooperation-with-hybrid-coe/253

[6] W. Steingartner, D. Moznik, and D. Galinec, "Disinformation campaigns and resilience in hybrid threats conceptual model," *2022 IEEE 16th International Scientific Conference on Informatics, Informatics 2022 - Proceedings*, pp. 287–292, 2022.

[7] D. J. Butts, S. A. Bollman, and M. S. Murillo, "Mathematical modeling of disinformation and effectiveness of mitigation policies," *Scientific Reports 2023 13:1*, vol. 13, pp. 1–12, 10 2023.

[8] M. Geers, B. Swire-Thompson, P. Lorenz-Spreen, S. M. Herzog, A. Kozyreva, and R. Hertwig, "The online misinformation engagement framework," *Current Opinion in Psychology*, vol. 55, p. 101739, 2024.

[9] F. Giglietto, G. Marino, R. Mincigrucci, and A. Stanziano, "A workflow to detect, monitor, and update lists of coordinated social media accounts across time: The case of the 2022 italian election," *Social Media and Society*, vol. 9, 2023.

[10] C. François, "Actors, behaviors, content: A disinformation abc highlighting three vectors of viral deception to guide industry & regulatory responses," 2019.

[11] S. Mirza, L. Begum, L. Niu, S. Pardo, A. Abouzied, P. Papotti, and C. Pöpper, "Tactics, threats & targets: Modeling disinformation and its mitigation," in *ISOC Network and Distributed Systems Security Symposium (NDSS)*, 2023.

[12] A. Bergh, "Understanding influence operations in social media: A cyber kill chain approach," *Journal of Information Warfare*, vol. 19, no. 4, pp. 110–131, 2020. [Online]. Available: https://www.jstor.org/stable/27033648

[13] A. Ahmad, J. Webb, K. C. Desouza, and J. Boorman, "Strategically-motivated advanced persistent threat: Definition, process, tactics and a disinformation model of counterattack," *Computers & Security*, vol. 86, pp. 402–418, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404818310988

[14] Ariane Bitoun, Antony Hubervic, Yann Prudent, "M&S for Influence Operations," NATO, Tech. Rep. STO-MP-MSG-149, 2017. [Online]. Available: https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-MSG-149/MP-MSG-149-07.pdf

[15] C. R. Walker, S.-J. Terp, P. C. Breuer, and C. L. Crooks, PhD, "Misinfosec," in *Companion Proceedings of The 2019 World Wide Web Conference*, ser. WWW '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1026–1032. [Online]. Available: https://doi.org/10.1145/3308560.3316742

[16] S. Terp and P. Breuer, "Disarm: a framework for analysis of disinformation campaigns," in *2022 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*, 2022, pp. 1–8.

[17] MITRE Corporation, "MITRE ATT&CK," 2013. [Online]. Available: https://attack.mitre.org

[18] Sep. 2020, [Online; accessed 27. Mar. 2024]. [Online]. Available: https://carnegieendowment.org/files/Pamment_-_Crafting_Disinformation_1.pdf

[19] K. C. Desouza, A. Ahmad, H. Naseer, and M. Sharma, "Weaponizing information systems for political disruption: The actor, lever, effects, and response taxonomy (alert)," *Computers & Security*, vol. 88, p. 101606, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404819301579

[20] K. M. Carley, "Social cybersecurity: an emerging science," *Computational and Mathematical Organization Theory*, vol. 26, pp. 365–381, 2020. [Online]. Available: https://doi.org/10.1007/s10588-020-09322-9

[21] "Scotch: A framework for rapidly assessing influence operations - atlantic council," https://www.atlanticcouncil.org/blogs/geotech-cues/scotch-a-framework-for-rapidly-assessing-influence-operations/, (Accessed on 03/27/2024).

[22] U.S Department of State, "Wagner Group, Yevgeniy Prigozhin, and Russia's Disinformation in Africa," 2022. [Online]. Available: https://www.state.gov/disarming-disinformation/wagner-group-yevgeniy-prigozhin-and-russias-disinformation-in-africa/

[23] J. Le Roux, Digital Forensic Research Lab (DFRLab), "Pro-russian facebook assets in mali coordinated support for wagner group, anti-democracy protests," 2022. [Online]. Available: https://medium.com/dfrlab/pro-russian-facebook-assets-in-mali-coordinated-support-for-wagner-group-anti-democracy-protests-2abaac4d87c4