ADAC.iO
Attribution Data Analysis Countermeasures Interoperability

THE DEVELOPMENT OF THE DISARM RED FRAMEWORK DURING ADAC.IO PROJECT

Summary of Activities – Final Report

ADAC.IO Publication – Deliverable 2.2
February 2024 – March 2026

Victoria Smith, Alliance4Europe
Adam Maunder, Alliance4Europe

Contact: victoria.smith@disarm.foundation | adacio.eu

Co-funded by the European Union

About ADAC.io: Attribution, Data, Analysis, Countermeasures and Interoperability

ADAC.io is a Horizon project funded by the European Union and coordinated by the Psychological Defence Research Institute at Lund University. It engages seven partners and has a three-year duration ranging from February 1, 2024 to January 31, 2027.

Based on the concept of Foreign Information Manipulation & Interference (FIMI) as elaborated by the EU EEAS, the purpose of ADAC.io is to protect democracy in the EU by strengthening the ability to deny the intended effects of FIMI on society. ADAC.io hence aims to significantly develop upon current knowledge of how FIMI can be detected, categorised, analysed, shared, and countered.

The project engages the following partners: Alliance4Europe (DE), Debunk EU (LT), Dortmund University - Institution of Journalism (DE), Cardiff University - Security, Crime and Intelligence Innovation Institute (UK); University of Social Sciences and Humanities (PL), Leiden University - The Hague Program for Cyber Norms (NL), Lund University - Psychological Defence Research Institute (SE).

This work was funded by the European Union Horizon Europe research and innovation program [grant number 101132444 – ADAC.io] and the UKRI under the UK government's Horizon Europe funding guarantee [grant number 10105669]. Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union, the European Commission or UKRI. Neither the European Union, the European Commission, nor the UKRI can be held responsible for them.

Summary of activities undertaken to support the development of the DISARM Red Framework and related technical standards during the Horizon-Europe funded ADAC.io Project

Final Report

Introduction

The DISARM Foundation supports the community working to detect, analyse and defend against Foreign Information Manipulation and Interference (FIMI) by maintaining and enhancing our publicly available frameworks and tools, such as the DISARM Red Framework - a taxonomy for classifying techniques associated with information threats, and our Microsoft Word Plugin tool, to support the tagging of reports with DISARM techniques.

Between February 2024 and March 2026, the Horizon-Europe funded ADAC.io project funded a series of activities to enhance and update the DISARM Red Framework, its associated tools, and related technical standards. This project has significantly accelerated the team's ability to gather and respond to feedback from users, understand the evolving needs of the community working to better understand threats to the information environment and release a series of framework updates and standards proposals that take these needs into consideration.

The result is an extensive reimagining of the DISARM Red Framework and the wider standards ecosystem, and the creation and improvement of tools and resources to support its use. We have tried to find ways to balance both the need for simplicity, and the need to support our users in faithfully describing the nuance and complexity of the information environments they study. This has not been easy. We have had to introduce more complexity in the early phases of this project, to remove subjectivity and inference from techniques, and make our techniques ready for AI and LLM assisted tools. We have done this in the hope that these new technologies, and future updates to the user interface will ultimately reduce complexity for the user.

The process of maintaining and enhancing the DISARM Framework is never finished, and work will continue beyond the funding this project has provided. ADAC.io has provided the first multi-year funding stream dedicated to enhancing the framework and its close relative, STIX. As such, it has allowed a small team to fundamentally reassess what the future of the DISARM Red Framework should look like, based on user feedback, assessments of evolutions in threat actor behaviour and the rapid acceleration in both the public adoption of AI and improvements in the underlying technologies.

This project has laid the foundations for our continued evolution of DISARM, for many years to come. We believe that the introduction of more objective, observable, techniques during the ADAC.io project will help to improve inter-annotator consistency, thereby enabling analysts around the world to quickly and efficiently share their work and objectively justify their findings. We hope this improves the capabilities of a wide range of DISARM users; helping shape the development of evidence-based policy changes, and building the case for law enforcement, sanctions and other defensive responses, to helping better communicate

threats to the information environment to the public or share information more effectively between partners and allies.

While there will always be more work to be done, ADAC.io has helped us shape the vision and develop a roadmap that extends well beyond the conclusion of this project. This report summarizes the work conducted to date, and our plans for future development and ends with a series of 14 annexes detailing the updates made to the DISARM Framework during the project and the materials created to support analysts who will use the new framework.

Project Outputs

Comprehensive Review of the DISARM Framework

Our first activity was an extensive outreach to the user community, to better understand their work, how they use the DISARM Framework, and how we could improve their user experiences of the Framework. We published the findings of this outreach as '*A Comprehensive Review of DISARM Framework and its Compatibility with Related Frameworks Used to Model Foreign Information Manipulation and Interference*'. This document laid out the challenges ahead of us, formed the basis for the activities conducted during the ADAC.io project, and should be regarded as a companion to this report.

The Comprehensive Review engaged a range of DISARM users and stakeholders, from governments to civil society organisations. It found that while some users wanted simplicity and fewer techniques, specific to their areas of work, others wanted techniques to be more objective and asked for techniques to be added to fill gaps in the framework. The Kill Chain structure, which one-to-one mapped individual techniques to Kill Chain stages, was considered confusing, and not in line with the analytical process followed by many of our users. Users wanted tools to support the laborious process of tagging reports with DISARM techniques and were interested in how DISARM could work more effectively alongside existing frameworks such as the Actor, Behaviour, Content, Distribution, Effect (ABCDE) model, or Meta's Online Operations Kill Chain.

In August 2025, we published a blogpost titled '*DISARM Red Retrospective and Futrospective*'.[1] This was a reflection of the findings of the comprehensive review of DISARM, conducted the previous year, and an explanation of how this review informed our approach to the development of the framework, and its tools and supporting resources to improve user experience. DISARM Red Framework updates in version 1.5 changed the way in which we added new techniques. We began introducing techniques that described one thing at a time and began to replace techniques that relied on users inferring behaviours with techniques that were observable. This allowed us to begin filling gaps in the framework while minimising the number of newly added techniques.

However, this approach created some complexity; some of the techniques we replaced were familiar to many users, and made it more difficult for users who preferred the broader, sensemaking techniques to navigate the framework. This lead us to introduce the idea that the DISARM Red Framework could consist of two parts; an observations framework that was more objective, and based on what the analyst could see, and an assessment framework, that allowed the analyst to use their observations in conjunction with other forms of intelligence (such as geopolitical or narrative analysis), to make assessments about the nature, scale or target of a FIMI incident or campaign.

[1] Annex 4, DISARM Red Retrospective and Futrospective

Implementing feedback in DISARM Red Framework updates

As previously stated, throughout the ADAC.io project we have tried to balance two competing areas of feedback; that DISARM needs to be more comprehensive and objective (to enable inter-annotator consistency and reflect the nuances of the information environment), and be simpler for analysts to use. This section reviews the series of updates made during the project in more detail.

We began to address this feedback with our v1.4, v1.5 and v1.6 updates,[2] released in March, August and November 2024 respectively. These updates updated, renamed, or replaced, techniques that were difficult to consistently apply. They also added new techniques, sub techniques and incidents.

Version 1.4 was a small update that addressed some pre-existing user requests. It saw 11 new techniques and 13 updated techniques, while two techniques were removed. This update occurred while we were conducting the Comprehensive Evaluation of DISARM.

Version 1.5 marked a significant milestone in the development of DISARM's Red Framework, and was the first update to respond to the findings of the Comprehensive Review. It introduced 38 new granular, observable techniques and sub techniques, allowing analysts to describe one thing at a time. It also found a way to fill gaps more efficiently and clarified definitions. Key features included;

- improvements to how personas could be catalogued;
- enhanced operability with Meta's Online Operations Kill Chain;
- the ability for analysts to document and investigate personas separately from making judgements about their legitimacy;
- Expanding persona documentation to allow for impersonated, fictional, parody and authentic personas; and
- Introducing a database of over 100 examples of DISARM Red techniques used in FIMI campaigns, helping analysts better understand how to apply them.

However, these updates increased the size and complexity of the framework, and the modification of some commonly used techniques caused some confusion for users. The update also introduced the idea of combining techniques, to describe behaviours in more detail, but DISARM's existing tools, such as the Navigator and Plug-In for Microsoft Word did not easily enable this workflow.

Version 1.6 added 82 techniques and sub techniques and over 200 examples of techniques used in reporting. The update focused on improving analysts' ability to document assets used by FIMI threat actors. This update addressed concerns expressed by analysts about their ability to use techniques such as *Create Inauthentic Accounts* and *Create Inauthentic Social Media Pages*, because of differences in interpretation of the term 'inauthentic'. It expanded the range of assets that could be documented (such as account, or online community group) and the types of platforms (eg, social media platform or microblogging platform) on which these assets were operating.

---

[2] Annex 1 - DISARM Version 1.4 Release Notes
Annex 2 - DISARM Version 1.5 Release Notes
Annex 3 - DISARM Version 1.6 Release Notes

These updates also supported ADAC.io project aims of improving DISARM's compatibility with other frameworks; v1.5 for example added techniques to improve interoperability with Meta's Online Operation Kill Chain.

While the v1.4, v1.5 and v1.6 updates addressed user feedback; filling gaps in the framework and clarifying technique names and descriptions, they also prompted a wider team reflection on the future direction of the DISARM Red Framework. It became clear that to give our users the best experience when using our framework, we would need to rethink both the methodology of the techniques we include, how we describe and categorise them, as well as the tools and interface for users to interact with and use them.

Planning the future of the DISARM Red Framework

The v1.4, v1.5 and v1.6 updates closed gaps in the framework, removed inferred elements from some techniques (such as the term 'inauthentic'), and clarified definitions. However, our approach also resulted in some challenges, and we learnt lessons along the way.

We found that the short technique names favoured in earlier versions in DISARM, created confusion and that analysts preferred longer, more descriptive technique names. Longer technique names mean that analysts do not necessarily have to click through to read a technique definition to clarify its meaning, saving time.

We found that through progressive updates we were adding many more techniques, and we needed to rethink how techniques were ordered and structured to ensure that analysts could easily find what they were looking for.

We learnt that we needed to deprecate techniques, rather than alter their names or definitions; users raised concerns about techniques that left room for interpretation being amended to become more specific, as this could create problems with previously tagged reports if the author had not interpreted the technique in the same way as the new definition.

We also found that some users who wanted a smaller framework of essential behaviours saw the kinds of techniques they had memorised being removed, and replaced with more granular, observable techniques. These users are least likely to want to learn a new system, or to trawl through documentation, so they were particularly impacted by these changes. This is also true of those who used the Framework for sensemaking, who liked browsing sense making techniques to help them think about the kinds of behaviours they might see in a campaign.

In our blog post '*DISARM Red Retrospective and Futrospective*[3]' we shared our reflections on what we felt had worked in previous updates, and what issues remained to be addressed. We also introduced the idea of the DISARM Red v2.0 Framework, which would allow us to separate different aspects of the current Red Framework, allowing us to reshape it so it was better suited to needs expressed by users.

Our strategy for delivering the DISARM Red v2.0 Framework was built around three pillars. First, we would explore ways to improve our internal process of delivering updates. We felt that by introducing *themed collections,* for example updates that focused on a particular user type or topic, we could better engage with specialists in particular communities of practice to deliver targeted updates more effectively. One example of this strategy is the v1.7 update,[4] released in January 2026, which engaged members of the fact checking community. This update focused on content verifiability documentation, suspicious features present in content, ways content can be reformatted and rhetorical devices. We included techniques describing evidence of falsified content or editing. As part of this update we also included techniques that enable analysts to better describe technology facilitated gender based violence.

The second pillar is focused on improving the resources that we provide to our users, to help them use the framework more efficiently and effectively. Examples here include the creation of the first DISARM playbooks[5] – reference guides for how to recognize and tag commonly

[3] Annex 4 - DISARM Red Retrospective and Futrospective
[4] Annex 14 - DISARM Version 1.7 Release Notes
[5] Annex 5 - DISARM Playbooks Overview

seen behaviours for specific threats or user workflows with procedures. Procedures are combinations of two or more observations to describe particular aspects of threat actor behaviour in more detail. So far, we have published support for three areas of key techniques;

- Technology-facilitated Gender-Based Violence (for DISARM v2),[6]
- Portal Kombat (for DISARM v2);[7] and
- Factchecking (for DISARM v1.7)[8]

We have also published guidance[9] on how to use the new observations framework to combine observations to create your own procedures. The ability to combine observations is the difference between, for example, making an observation that you have seen a user 'create a post', to saying that you have seen a 'News article falsely attributed to a real journalist' (DISARM Tags: *(T0156.001: Create Post (T0171.000: Text Content, T0176.001: News Report, T0178.002: Content Presented as Produced by Third Party (T0097.102: Journalist Persona (T0143.003: Impersonated Persona)), T0178.001: Incorrect Content Source Presented))*). The latter provides much more information about the type of post that has been created and the context of the account that posted it.

We recognise that procedures create a new level of complexity for DISARM and we are exploring ways to mitigate this, both by producing resources, such as playbooks, that pre-write procedures for commonly seen behaviours, and by exploring how AI, LLMs, an improved user interface and other technologies can provide more support.

We hope to gradually move towards a system where an analyst can see a human-readable description of a procedure e.g. 'News articles falsely attributed to real journalists', and the machine handles the observable techniques that lie behind it (in this case, seven). These individual techniques can be revealed to an analyst wishing to conduct further, in depth research, but are simplified for non-technical audiences such as policy makers and senior leaders, without compromising on the detail and evidence collection behind them.

Finally, we have worked closely with our user community to understand how to make improvements to the framework. We are deprecating, rather than removing techniques, to ensure reports tagged with previous versions of DISARM don't become obsolete. We write extensive release notes for each new DISARM version that explain why we have deprecated the techniques we have, and how these behaviours can be tagged with the updated framework. The idea is to enable both backwards compatibility with reports tagged with earlier versions, but also the potential for more effective updating, should a user wish to do so.

---

[6] Annex 6 - DISARM TFGBV Playbook - Key Techniques
Annex 7 - DISARM TFGBV Playbook Tagged Reports
Annex 8 - DISARM TFGBV Playbook - Prefabricated Procedures

[7] Annex 13 - DISARM Portal Kombat Playbook - Tagging Support

[8] Annex 10 - DISARM Fact Checkers Playbook - Key Techniques
Annex 11 - DISARM Fact Checkers Playbook - Tagged Reports
Annex 12 - DISARM Fact Checkers Playbook - Tagging Support

[9] Annex 15 - DISARM v2.0 Actions Rules

Redesigning the DISARM Red Framework - Observations

The ADAC.io project has given DISARM the freedom to completely rethink its future, and we decided that to best achieve this, we could not be bound by previous thinking and methodology, if we felt it was not the right fit for our users. This project also began just as publicly available AI models were gathering pace. Our challenge is to build a framework that is flexible enough to accommodate changes in technology and how these technologies are used to manipulate the information environment, whether or not we can predict what these changes might be.

Through the process of thinking about different ways to structure the framework, we reflected on the layout of supermarkets; supermarkets can be very big, they can expand product lines and sell anything from homewares to food, or clothes to postage stamps. However, we understand the shared system that supermarkets operate within – we know the fruit and vegetables are placed together, there are sections for raw meat and cooked foods, tinned goods, home baking, beverages etc. We can walk into a supermarket of any size and navigate around it with relative ease, avoiding the aisle for baby food if we don't have young children and heading straight for the fresh fruit and vegetables if that is what we need. We needed to replicate a similar system for DISARM, allowing it to expand to accommodate techniques of the future, without increasing the complexity for the user.

To do this, it was important to make a clear difference between observations of what can be seen, and what can be inferred from these observations. Analytical inference is essential to understanding the information environment; no analyst can stand behind a threat actor and observe everything they do, some things must be inferred in order to make sense of the nature, scale, impact or target audience of a FIMI incident. We therefore decided to separate the framework into two parts; Observation and Assessment. During this process, it was clear to us that the DISARM v1 layout, centred on a kill chain in which each link represents a specific goal of the adversary, was not appropriate. Individual techniques do not easily map one-to-one with kill chain stages and what stage an influence operation is at is often an assessment dependent on a range of criteria.

Because it is not easy to one-to-one map observable techniques to the kill chain, many users found it difficult to find the techniques they wanted; perhaps they were looking in the wrong column and assumed it wasn't there, or perhaps when they saw it in a column that they did not feel mapped well to where they thought it should be, they assumed that it was therefore not appropriate for them to use in that context. The kill chain running left to right also put many of the aspects of an operation that are hardest to identify right at the 'start', for example strategic and tactical objectives. This was discouraging for many users, and particularly confusing for new users.

We therefore decided that the DISARM Red v2.0 framework would operate in a different way. Our users reported being familiar with the Actor, Behaviour, Content, Distribution, Effect (ABCDE) model, so we decided to replicate a similar structure, but beginning with the A, B and C, creating an Asset, Action, Content matrix. We felt this would better support alignment with the pre-existing ABCDE model, while also making it easier for analysts to find the techniques they needed.

We reflected that as 'Actor' implies attribution, which has its own complexities and is not a prerequisite for beginning an investigation, our 'A' would stand for Asset. This better

reflected the techniques in our framework designed to help analysts describe how an asset, for example a social media account, is presenting itself online, without having to make an assessment about whether or to what extent that account is a real human presenting their real identity.

While we are expanding the scope of techniques we include, in the past DISARM has been seen primarily as a 'behaviour' framework. We therefore thought it would be better for our second category to be 'Actions'. Our thinking was that we needed to help our users describe the assets they observe (whether they can attribute them or not), the actions taken by these assets, and the types of content these assets create. We believe that with this framework it will be much more intuitive for an analyst to work their way around the framework, avoiding the 'aisles' they don't need and quickly finding those they do.

We have published a beta version of the DISARM Red v2.0 Observation Framework here.

Redesigning the DISARM Red Framework – Assessments

The DISARM Red version 1.0 Frameworks include techniques that are often inferred such as those in the 'Plan Strategy' column, or that mix inferred and observed elements, for example 'create fake account'. Our users have told us that while they recognise that some of our techniques are subjective, the ability to make evidence-based assessments, which often requires a degree of subjectivity or inference, is important to them.

To this end, we have developed the first draft of an Assessment Framework[10] to support the Observations Framework. This Assessment Framework is designed to help communicate the context of the threat and the suspected threat actor, for example the nature and scale of the campaign, who it is targeting and the objective and capabilities of the threat actor.

The ability to assess techniques in their wider strategic and operational context is an important step in conveying the threat to a wider audience, from policy makers to the general public. Knowing that an account is presenting as a journalist, publishing news about where displaced people in a war zone can access basic supplies would take on a whole new dimension if it turned out the journalist persona was fake and was being used by a threat actor to lure civilians into danger with the promise of aid supplies.

The Assessment Framework can help analysts build contextual and behavioural evidence for an attribution, determine how far a campaign has spread, who it appears to be targeting, and what the motivations of the campaign or threat actor might be.

Developing our toolset:

DISARM Add-In for Microsoft Word
ADAC.io funding has contributed to the development of the DISARM Add-In for Microsoft Word, known colloquially as the 'tagger tool'. This tool helps streamline the manual tagging process for analysts as they write their reports. As an official Add-In written in JavaScript and available in the Microsoft Store, it does so securely by running within an isolated process with no access to the local filesystem. It replaces the DISARM Plug-In for Microsoft Word, which was written in Visual Basic for Applications and which is no longer supported.

With ADAC.io funding we now offer a version of the DISARM Add-In for annotating reports using the DISARM 2.0 Observations Framework, with support for Word on MacOS, for Word on Windows, and for Word on the Web. An auto-complete feature expedites technique selection, analysts can now tag entities and observables as well as behaviors, and annotated reports can now be viewed using the DISARM Navigator or knowledge graphs built automatically using Neo4J. Future aspirations include smart search, a chatbot interface, nested tagging, and support for alternative text editors such as Google Docs and Libre Office.

DISARM Annotator Tool
We are also developing an AI-assisted annotation tool, which leverages AI model inference to analyse uploaded reports and generate contextually relevant annotation suggestions. Users can review, accept, or dismiss these suggestions, maintaining human oversight throughout the annotation process. The tool supports the STIX standard for both import and export: existing STIX bundles can be imported to resume or enrich partially completed annotations, while

[10] Annex 16 - DISARM v2.0 Assessment Framework

finished reports can be exported as STIX bundles for direct ingestion into OpenCTI. As part of the ADAC.io project we are making a prototype of this tool available for a small number of invited users to enable testing.

To give users flexibility and control, the tool is designed with a multi-provider architecture, enabling users to select their preferred AI model provider and authenticate via their own API keys. This approach transfers inference costs directly to the user, reducing reliance on DISARM to fund AI queries. The tool will also incorporate structured feedback mechanisms, allowing users to make recommendations that can inform the ongoing development of both the annotation tool and the DISARM framework.

Another future aspiration is for DISARM to expand its pool of tagged reports to assist with the development of DISARM training, improve the quality of the AI model outputs and enable analysis of an increasingly large pool of data to identify trends (to inform the development of resources such as playbooks) or gaps in the framework (to assist the development of new techniques). The annotator tool will be a key enabler of this, as it will significantly increase the speed at which reports can be tagged.

User Interface

Developments to the user interface are beyond the scope of the ADAC.io project. However, the move to the DISARM Red v2.0 Framework developed during this project has exposed limitations in the Navigator tool, which we are currently using. Beyond this project, the DISARM Foundation will continue to explore ways to improve our user experience, which will likely include changes to the user interface.

In the meantime, we will be introducing multi version support in the Navigator, to allow users to view different versions of the framework. We are also building an OpenCTI connector to allow users to find and use the beta version of the v2.0 Observations Framework in OpenCTI.

These changes will be important to help us reduce complexity for our users, one way would be the ability for a user to apply filters to the framework so they can focus on specific research questions, for example ‘I want to describe how an account presents itself online’, which might focus on persona types, account imagery etc, or ‘I want to conduct a factcheck’, which would bring back very different techniques, focusing on how to describe the types of content being investigated.

Focusing on research questions will also help us align the DISARM Red v2.0 and attribution frameworks, in line with the aspiration of the grant agreement. Just as it can be difficult to conduct one-to-one mapping of techniques to one kill chain stage, it can also be difficult to one-to-one map techniques to the four categories in the attribution framework; some techniques could apply to multiple attribution categories depending on the context. However, focusing on a research area can also help analysts understand how the information they are looking at could be used to support an attribution. For example, ‘I want to describe how an account presents itself online’ could provide contextual evidence, ‘I want to describe the actions taken by an account’ could provide behavioural evidence, and ‘I want to understand how an asset solicits funds from others’ could provide technical (financial) evidence.

While we continue to explore ways to introduce research question decision trees to the DISARM tool set, we believe that the reorganisation of techniques in the v2.0 Observations Framework will improve alignment with the attribution framework, and support analysts who

wish to make attributions, by helping analysts to focus on documenting evidence relating to the asset (supporting, for example, contextual or behavioural evidence), Actions (supporting behavioural or technical evidence), and Content (supporting contextual evidence).

Extending Existing Technical Standards to Support FIMI – Improving Interoperability

One of the goals of the ADAC.io project has been to work with the standards body, OASIS, to propose extensions to the existing standard for modeling cyber threats which will enable the comprehensive modeling of FIMI and hybrid threats and the sharing of threat intelligence among members of Information Sharing and Analysis Organisations such as the FIMI-ISAC. Since the start of the project, members of the ADAC.io team have worked with OASIS on the DAD-CDM project to create a Technical Steering Committee to guide the community in defining objects, relationships, and taxonomies required to extend the STIX 2.1 standard for FIMI. The resulting Common Data Model will empower humans and machines to model threats to the information environment at speed and scale.

In February 2025 the Technical Steering Committee published a summary of their key findings and progress. More specifically, members of ADAC.io have crafted draft proposals that detail extensions to the Threat Actor subsystem based on a revision of the original work carried out by Intel Corporation, and extensions to the Incident subsystem proposed by the Cyber Threat Intelligence Technical Committee for consideration in STIX 2.2. The Technical Steering Committee is working on additional proposals for the Narrative and Channel subsystems. After gathering feedback the extensions will be finalised and made available in the Common Object Repository.

Beyond ADAC.io, the future of the DISARM Framework

As previously stated, the DISARM v2 Red Framework marks a step change in direction for DISARM. The developments achieved in this project mark the start of a series of planned changes.

First, the redesign of the Red Framework is the first step towards setting DISARM in a wider analysis, intelligence and response cycle. The Observations Framework provides evidence for specific techniques used by threat actors. The new design helps the analyst piece these techniques together, in ways that provide a richer, deeper context of the observable activity. This provides evidence to those wishing to take this work a step further and make assessments.

Assessments are both a communication tool to a wider, non-technical audience and a way of providing more context to inform a defender's understanding of the threat as well as their options to respond to it or defend against it.

The developments to the DISARM Red v2.0 Framework are contributing to a wider redesign of DISARM's workflow and technology stack which we hope will improve user experience, and facilitate the creation of a bulk standardized data set that can be easily shared and analysed.

We have designed a twelve-step analyst workflow, to support an analyst's use of DISARM in the analysis cycle. We hope this helps our users better understand where DISARM can support their analysis and assessment work. A summary of the planned workflow is as follows:

Phase 1: Identification and ingestion of data
Step 1 – FIMI Candidate Declared: An incident is flagged via technology, reports, or threat intelligence. It is defined as a 'FIMI Incident' - a describable and temporally finite act of possible foreign information manipulation or interference.

Step 2 – Standardize Input: Data (including: written reports, structured reporting, behavioural data, etc.) is converted into a structured FIMI incident report, known as a FIMI Incident Pack. This pack comprises information such as a summarised description of the FIMI incident; a series of manual fields already inputted; a table of behavioural evidence directly constituting the FIMI incident; and a series of automated annotations that have already been inputted.

Phase 2: Technical analysis
Step 3 – Entity Recognition: Systematic identification of directly observable objects that can be identified in any FIMI Incident Pack such as people, places, and organizations, added as structured annotations.

Step 4 – Observable TTPs: Uses the DISARM v2 Red framework to tag techniques visible in the data. New, unrecognized techniques may also be identified for future evaluation.

Step 5 – Capability and Infrastructure Assessment: Evaluates the technical "back-end," such as the actor's budget, team sophistication, and underlying digital infrastructure, by assessing

information collected about the incident and any related external data leaks or media reporting.

Phase 3: Strategic assessment and attribution
Step 6 – Tactical Objectives: Assesses tactical objectives, such as intent regarding the FIMI's narrative manipulation or network manipulation. The assessments are codified using the DISARM Assessments Framework and added to the FIMI Pack as additional annotations.

Step 7 – Attribution: Assesses responsibility using the NATO Hybrid CoE framework, which provides a way of identifying behavioural, contextual and technical attribution signals, alongside the navigation of legal and ethical considerations when making attributions public. It acknowledges different levels of confidence, and that attribution can occur at different levels of detail, from broad actor types to specific individuals.

Step 8 – Contextualisation: Moves from looking at the incident in isolation to linking it with other reports. This includes identifying overarching campaigns, kill chain timing, and shared technical assets.

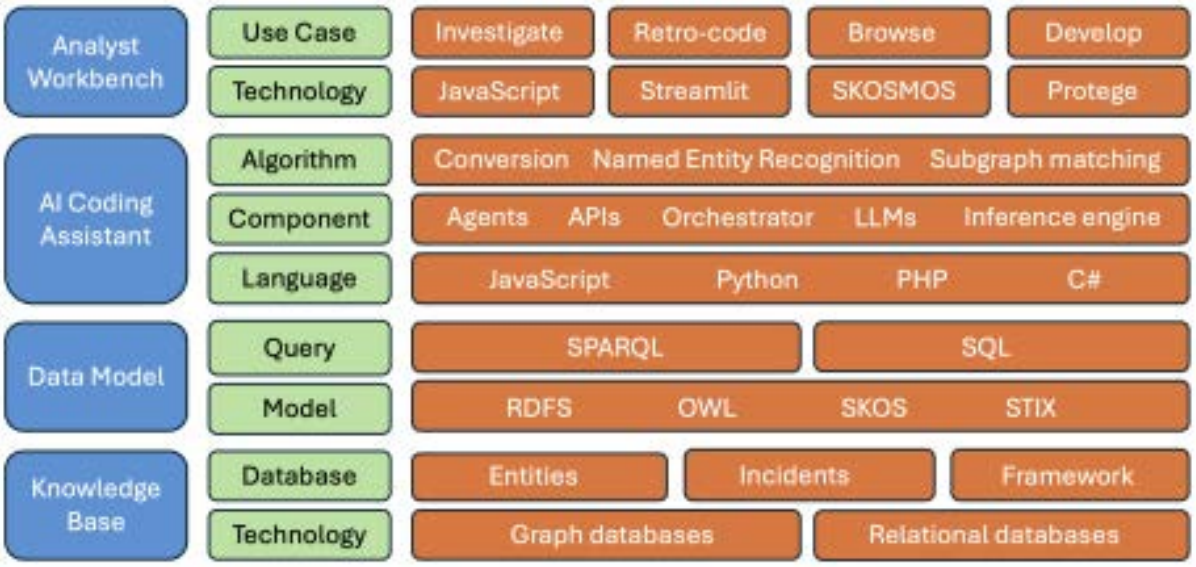
Phase 4: Impact and strategic response
Step 9 – Target Audience and Impact Assessment: Uses the DISARM Assessment Framework to assess the intended audience. It measures the success of the incident through analytics, polling, the "Breakout Scale", the "Risk-Impact Index" or other forms of evaluation. Step 9 concludes with two forms of annotation added to the FIMI Pack: (a) an assessment of the communities or groups targeted by the FIMI Incident, and the vulnerabilities (if present) that were likely exploited, and (b) an impact assessment of the emotional, social, institutional and financial impact of the FIMI incident.

Step 10 – Strategic Objectives: Uses the DISARM Assessments Framework to help an analyst determine the strategic objective of the FIMI incident.

Step 11 – Response: Uses the D-RAIL methodology to help an analyst to develop mitigation strategies. This is a complex step that involves mapping actor vulnerabilities, generating counter-responses, and managing the risks and costs of the intervention.

Step 12 – Output: The final step involves the creation of a standardised output that contains all of the outputs created by each of the stages described in this report in a structured form. The outputs can be brought together into a knowledge graph, linking various incidents and assessments to enable long-term intelligence.

To achieve this, we have developed a plan to redesign the DISARM technical architecture. While this is beyond the scope of the ADAC.io project to deliver, it will be grounded in the lessons we have learned and the developments we have been able to achieve during this project, and will ensure that ADAC.io project deliverables continue to deliver benefit to our users for years to come. We envision a state-of-the-art technology stack that brings together a sophisticated web interface with an agentic AI architecture and a knowledge organization system built using a FIMI ontology and semantic reasoning engine. As outlined below, the technology stack is composed of four principal parts: an Analyst Workbench, an AI layer, a common data model, and an underlying knowledge base. The image below provides an overview of the plan for this technical architecture.



Proposed technical architecture for DISARM 2.0

DISARM Foundation Resources

The latest DISARM resources are linked from the DISARM website: https://www.disarm.foundation/framework

Framework and Tools

DISARM v2 Observations Framework (Beta):
https://github.com/DISARMFoundation/DISARMframeworks-20-observable

DISARM Navigator for v2 Observations Framework (Beta):
https://disarmfoundation.github.io/disarm-navigator-20-observable/

DISARM STIX Bundle for v2 Observations Framework (Beta):
https://github.com/DISARMFoundation/disarm-stix-20-observable

DISARM Navigator for v1.7: https://disarmfoundation.github.io/disarm-navigator-17/

DISARM v1.6 Add-In for Microsoft Word:
https://marketplace.microsoft.com/en-us/product/saas/wa200008045?tab=overview

DISARM v2.0 Add-In for Microsoft Word:
https://github.com/DISARMFoundation/DISARMframeworks-20-observable/blob/main/CODE/manifest2.0.xml

Installation Instructions for DISARM v2.0 Add-In for Microsoft Word:
For Word on MacOS:
https://learn.microsoft.com/en-us/office/dev/add-ins/testing/sideload-an-office-add-in-on-mac
For Word on Windows:
https://learn.microsoft.com/en-us/office/dev/add-ins/testing/create-a-network-shared-folder-catalog-for-task-pane-and-content-add-ins
For Word on the Web:
https://learn.microsoft.com/en-us/office/dev/add-ins/testing/sideload-office-add-ins-for-testing

DISARM v2.1 Observations Framework (Beta, Excel only),
https://docs.google.com/spreadsheets/d/14LxE_0woxhgpTf3NousqSpP5nOso2aIY/edit?gid=1430161519#gid=1430161519

Initial Exploration of Threat Assessments, March 2026,
https://docs.google.com/document/d/1eUwDFM13yKBLZUoxbJa4tSxvi4YLZFXu2RBGh8MAz0U/edit?tab=t.0#heading=h.uq1nnhy9jder

Assessments Framework v2.1 (Beta, Excel Only), March 2026,
https://docs.google.com/spreadsheets/d/1m24dIBjAQC4t6_0gm_9ELmSHtc4g3ti3V9oCxQgvMbA/edit?gid=523500252#gid=523500252

Release Notes

DISARM v1.6: Non-Content Assets Update (November 2024): https://medium.com/disarming-disinformation/disarm-v1-6-non-content-assets-update-f627692f5c0d

DISARM v1.5: Personas Update (August 2024): https://medium.com/disarming-disinformation/disarm-v1-5-personas-update-bf0323614e3b

DISARM Update: Version 1.4 (March 2024): https://medium.com/disarming-disinformation/disarm-update-version-1-4-b0b2ea867d0b

DISARM Update: Version 1.3 (September 2023): https://medium.com/disarming-disinformation/disarm-update-version-1-3-9dfcf2a29864

Resources

DISARM Red Retrospective and Futrospective (August 2025): https://medium.com/disarming-disinformation/disarm-red-retrospective-and-futrospective-68854e51c5cf

A Comprehensive Review of DISARM Framework and its Compatibility with Related Frameworks Used to Model Foreign Information Manipulation and Interference (August 2024): https://adacio.eu/disarm-august2024

DISARM Progress Update: DISARM v2, 4 March 2026, https://medium.com/disarming-disinformation/disarm-progress-update-disarm-v2-64dfa62ca97c

Actions Rules, Rules for creating procedures in the Observations Framework, 3 March 2026, https://docs.google.com/document/d/1fKuv6N82NV75bT4YMx7nrtuqPQ9lXo5eh-Aahl32O6c/edit?tab=t.0#heading=h.76ww6ypwzppz

DISARM Foundation GitHub page: https://github.com/disarmfoundation

DISARM Playbooks, 13 March 2026, https://medium.com/disarming-disinformation/disarm-playbooks-180cf59cdc7f

Proposal to Extend STIX to Support FIMI

Summary of Key Findings: https://github.com/DAD-CDM/dad-cdm-tsc/blob/main/DAD-CDM-Key-Findings-202502.md

Extending the Threat Actor subsystem: https://drive.google.com/file/d/1DOqcBkEyThYtYaiTvQsNeXteko7qmq2l/view?usp=drive_link

Extending the Incident subsystem: https://drive.google.com/file/d/1odz1lPVbPqGFRibZIq9wmUiWVdsf5yGg/view?usp=drive_link

Annex 1 - DISARM Version 1.4 Release Notes

DISARM Version 1.4

DISARM update version 1.4 has been produced as part of the Horizon Funded project ADAC.io; “Attribution – Data – Analysis – Countermeasures – Interoperability”. This initial update focuses on delivering based on existing user feedback.

Overview

Large Changes

T0085: Develop Text-Based Content

Added T0085.005: Develop Book

Added T0085.006: Develop Opinion Article

Updated T0085.002: Develop False or Altered Documents

Removed T0089.002: Create Inauthentic Documents

T0019: Generate Information Pollution

Moved T0019: Generate Information Pollution

Moved T0019.001: Create Fake Research

Removed T0019.002: Hijack Hashtags

Updated T0049: Flooding the Information Space

Updated T0049.002: Hijack Existing Hashtag

T0011: Compromise Legitimate Accounts

Added T0141: Acquire Compromised Asset

Updated & Moved T0011: Compromise Legitimate Accounts

Added T0141.002: Acquire Compromised Website

TA08: Conduct Pump Priming

Moved T0113: Employ Commercial Analytics Firms

Moved T0039: Bait Legitimate Influencers

T0099: Prepare Assets Impersonating Legitimate Entities

Updated T0099: Prepare Assets Impersonating Legitimate Entities

Updated T0099.001: Astroturfing

Added T0099.003: Impersonate Existing Organisation

Added T0099.004: Impersonate Existing Media Outlet

Added T0099.005: Impersonate Existing Official

Added T0099.006: Impersonate Existing Influencer

Small Changes

Updated T0097.001: Backstop Personas

Updated T0104.002: Dating Apps

Updated T0049: Flooding the Information Space

Updated TA15: Establish Social Assets

Updated TA05: Microtarget

Incidents

Added for T0104.002: Dating Apps

Added for T0141.001: Acquire Compromised Account

Added for T0141.002: Acquire Compromised Website

Request for Feedback

TA07: Select Channels and Affordances

Funded by the European Union

Large Changes

T0085: Develop Text-Based Content

Added T0085.005: Develop Book
Added T0085.006: Develop Opinion Article
Updated T0085.002: Develop False or Altered Documents
Removed T0089.002: Create Inauthentic Documents

We've introduced two new sub-techniques to T0085: Develop Text-Based Content, to allow tagging of different types of text-based content. Let us know which other types of text-based content you'd like to be able to tag!

Previous	Updated
None	**TA06: Develop Content** **T0085: Develop Text-Based Content** **T0085.005: Develop Book** Summary: Produce text content in the form of a book. This technique covers both e-books and physical books, however, the former is more easily deployed by threat actors given the lower cost to develop.
None	**TA06: Develop Content** **T0085: Develop Text-Based Content** **T0085.006: Develop Opinion Article** Summary: Opinion articles (aka "Op-Eds" or "Editorials") are articles or regular columns flagged as "opinion" posted to news sources, and can be contributed by people outside the organisation. Flagging articles as opinions allow news organisations to distinguish them from the typical expectations of objective news reporting while distancing the presented opinion from the organisation or its employees. The use of this technique is not by itself an indication of malicious or inauthentic content; Op-eds are a common format in media. However, threat actors exploit op-eds to, for example, submit opinion articles to local media to promote their narratives. Examples from the perspective of a news site involve publishing op-eds from perceived prestigious voices to give legitimacy to an inauthentic publication, or supporting causes by hosting op-eds from actors aligned with the organisation's goals.

T0085.003: Develop False or Altered Documents was also updated. Ideally we want Techniques to cover one unique behaviour, but this Technique had three potential implications;

1. Text was presented in the form of a Document
2. The document's text contained false information, and/or

Funded by the European Union

3. The document's text had been appropriated and altered from a previous legitimate source.

Going forward this Technique will only imply that threat actors delivered text in the form of a document under the name *T0085.004: Develop Document*.

By removing the requirement to assert that a document is false, analysts can focus on identifying the format that content was delivered in (rather than what it said).

The alteration of existing legitimate documents can be tagged using the existing Technique *T0089.003: Alter Authentic Documents* (which also had a small typo in its summary fixed).

To avoid causing issues with backwards compatibility, *Develop Document* was assigned a new ID, instead of keeping the ID used by *Develop False or Altered Documents*.

Previous	Updated
TA06: Develop Content **T0085: Develop Text-Based Content** **T0085.002: Develop False or Altered Documents** Summary: *None*	**TA06: Develop Content** **T0085: Develop Text-Based Content** **T0085.004: Develop Document** Summary: Produce text in the form of a document
TA06: Develop Content **T0089: Obtain Private Documents** **T0089.003: Alter Authentic Documents** Summary: Alter authentic documents (public or non-public) to achieve campaign goals. The altered documents are intended to appear as if they are authentic can be "leaked" during later stages in the operation.	**TA06: Develop Content** **T0089: Obtain Private Documents** **T0089.003: Alter Authentic Documents** Summary: Alter authentic documents (public or non-public) to achieve campaign goals. The altered documents are intended to appear as if they are authentic and can be "leaked" during later stages in the operation.

While making this change, we removed T0089.002: Create Inauthentic Documents.

The original design intent of this Technique was to document cases where defenders produce real-looking but inauthentic documents in places where a hacker is likely to steal them. This was a strategy employed by France in 2017; "a classic "cyber-blurring" strategy, well known to banks and corporations, creating false email accounts and filled them with phony documents the way a bank teller keeps fake bills in the cash drawer in case of a robbery."

As a defensive behaviour, this Technique doesn't fit the Red framework, and can easily be confused with other document-based Techniques, so it made sense to deprecate the Technique.

Previous	Updated
TA06: Develop Content **T0089: Obtain Private Documents** **T0089.002: Create Inauthentic Documents** Summary: Create inauthentic documents intended to appear as if they are authentic non-public documents. These documents can be "leaked" during later stages in the operation	*Removed*

Funded by the European Union

T0019: Generate Information Pollution

Moved T0019: Generate Information Pollution
Moved T0019.001: Create Fake Research
Removed T0019.002: Hijack Hashtags
Updated T0049: Flooding the Information Space
Updated T0049.002: Hijack Existing Hashtag

Generate Information Pollution has been updated to be a Subtechnique of T0049: Flooding the Information Space, which better suits this Technique's methods than its previous home of TA06: Develop Content. Its summary has also been updated for clarity.

Previous	Updated
TA06: Develop Content **T0019: Generate Information Pollution** Summary: Flood social channels; drive traffic/ engagement to all assets; create aura/sense/perception of pervasiveness/consensus (for or against or both simultaneously) of an issue or topic. "Nothing is true, but everything is possible." Akin to astroturfing campaign.	**TA17: Maximise Exposure** **T0049: Flood Information Space** **T0049.008: Generate Information Pollution** Summary: Information Pollution occurs when threat actors attempt to ruin a source of information by flooding it with lots of inauthentic or unreliable content, intending to make it harder for legitimate users to find the information they're looking for. This Subtechnique's objective is to reduce exposure to target information, rather than promoting exposure to campaign content, for which the parent Technique T0049 can be used. Analysts will need to infer what the motive for flooding an information space was when deciding whether to use T0049 or T0049.008 to tag a case when an information space is flooded. If such inference is not possible, default to T0049. This Technique previously used the ID T0019

Generate Information Pollution's sub-techniques are also being updated, and as part of this change T0019.002: Hijack Hashtags is being merged into T0049.002: Hijack Existing Hashtag.

These Subtechniques essentially covered the same behaviour (i.e. the flooding of a hashtag) for different motives (ruining hashtag functionality in the former, and maximising exposure to campaign content in the latter). By updating T0049.002 to cover both motives (and updating its name to Flood Existing Hashtag in accordance), we free analysts to simply tag the observable behaviour of flooding a hashtag without requiring them to infer motive first.

We considered instead introducing *Pollute Existing Hashtag* to allow analysts to tag the use of flooding a hashtag for the purpose of ruining a source of information when they are able to make this inference, but we decided this would risk causing confusion for not enough benefit.

Previous	Updated

Funded by the European Union

TA17: Maximise Exposure **T0049: Flooding the Information Space** **T0049.002: Hijack Existing Hashtag** Summary: Take over an existing hashtag to drive exposure	**TA17: Maximise Exposure** **T0049: Flood Information Space** **T0049.002: Flood Existing Hashtag** Hashtags can be used by communities to collate information they post about particular topics (such as their interests, or current events) and users can find communities to join by exploring hashtags they're interested in. Threat actors can flood an existing hashtag to try to ruin hashtag functionality, posting content unrelated to the hashtag alongside it, making it a less reliable source of relevant information. They may also try to flood existing hashtags with campaign content, with the intent of maximising exposure to users. This Technique covers cases where threat actors flood existing hashtags with campaign content. This Technique covers behaviours previously documented by T0019.002: Hijack Hashtags, which has since been deprecated. This Technique was previously called Hijack Existing Hashtag.
TA06: Develop Content **T0019: Generate Information Pollution** **T0019.002: Hijack Hashtags** Summary: Hashtag hijacking occurs when users "[use] a trending hashtag to promote topics that are substantially different from its recent context" (VanDam and Tan, 2016) or "to promote one's own social media agenda" (Darius and Stephany, 2019).	*Removed*

Create Fake Research has been updated to be a sub-technique of *Develop Text-Based Content.* With *Generate Information Pollution's* move to *Flooding the Information Space,* this felt like a good time to move *Create Fake Research* to a Technique which better encapsulates it, and gives our users the freedom to tag inauthentic research when it's used in non-polluting campaigns; threat Actors can use any type of content to pollute an information environment, and inauthentic research can be used in operations that aren't intending to pollute an information environment.

Previous	Updated
TA06: Develop Content **T0019: Generate Information Pollution** **T0019.001: Create Fake Research** Summary: Create fake academic research. Example: fake social science research is often aimed at hot-button social issues such as gender, race and sexuality. Fake science research can target Climate Science debate or pseudoscience like anti-vaxx	**TA06: Develop Content** **T0085: Develop Text-Based Content** **T0085.007: Create Fake Research** Summary: Create fake academic research. Example: fake social science research is often aimed at hot-button social issues such as gender, race and sexuality. Fake science research can target Climate Science debate or pseudoscience like anti-vaxx This Technique previously used the ID T0019.001

Funded by the European Union

T0011: Compromise Legitimate Accounts

Added T0141 Acquire Compromised Asset
Updated & Moved T0011: Compromise Legitimate Accounts
Added T0141.002: Acquire Compromised Website

Previously *T0011: Compromise Legitimate Accounts* was the only way to tag cases where threat actors compromised existing assets to distribute content.

We've introduced the new *T0141: Acquire Compromised Asset* to provide a broad Technique which can cover cases where things other than accounts are taken over by threat actors, along with specific sub-techniques for hacking into accounts and websites.

Previous	Updated
None	**TA15: Establish Assets** **T0141: Acquire Compromised Asset** Summary: Threat Actors may take over existing assets not owned by them through nefarious means, such as using technical exploits, hacking, purchasing compromised accounts from the dark web, or social engineering.
TA16: Establish Legitimacy **T0011: Compromise Legitimate Accounts** Summary: Hack or take over legitimate accounts to distribute misinformation or damaging content	**TA15: Establish Assets** **T0141: Acquire Compromised Asset** **T0141.001: Acquire Compromised Account** Summary: Threat Actors can take over existing users' accounts to distribute campaign content. The actor may maintain the asset's previous identity to capitalise on the perceived legitimacy its previous owner had cultivated. The actor may completely rebrand the account to exploit its existing reach, or relying on the account's history to avoid more stringent automated content moderation rules applied to new accounts. See also [Mitre ATT&CK's T1586 Compromise Accounts](https://attack.mitre.org/techniques/T1586/) for more technical information on how threat actors may achieve this objective. This Technique was previously called Compromise Legitimate Accounts, and used the ID T0011.

Funded by the European Union

None	**TA15: Establish Assets** **T0141: Acquire Compromised Asset** **T0141.001: Acquire Compromised Website** Summary: Threat Actors may take over existing websites to publish or amplify inauthentic narratives. This includes the defacement of websites, and cases where websites' personas are maintained to add credence to threat actors' narratives. See also [Mitre ATT&CK's T1584 Compromise Infrastructure](https://attack.mitre.org/techniques/T1584/) for more technical information on how threat actors may achieve this objective.

T0099: Prepare Assets Impersonating Legitimate Entities

Updated T0099: Prepare Assets Impersonating Legitimate Entities
Updated T0099.001: Astroturfing
Added T0099.003: Impersonate Existing Organisation
Added T0099.004: Impersonate Existing Media Outlet
Added T0099.005: Impersonate Existing Official
Added T0099.006: Impersonate Existing Influencer

We're introducing sub-techniques to T0099: Prepare Assets Impersonating Legitimate Entities, in order to allow tracking of which types of existing entities are being impersonated. We're also renaming T0099 to Impersonate Existing Entity, which is shorter, and doesn't require a value judgement on what "legitimate" entities are.

Previous	Updated
TA16: Establish Legitimacy **T0099: Prepare Assets Impersonating Legitimate Entities** Summary: An influence operation may prepare assets impersonating legitimate entities to further conceal its network identity and add a layer of legitimacy to its operation content. Users will more likely believe and less likely fact-check news from recognisable sources rather than unknown sites. Legitimate entities may include authentic news outlets, public figures, organisations, or state entities. An influence operation may use a wide variety of cyber techniques to impersonate a legitimate entity's website or social media account. Typosquatting87 is the international registration of a domain name with purposeful variations of the impersonated domain name through intentional typos, top-level domain (TLD) manipulation, or punycode. Typosquatting facilitates the creation of falsified websites by creating similar domain names in the URL box, leaving it to the user to confirm that the URL is correct.	**TA16: Establish Legitimacy** **T0099: Impersonate Existing Entity** Summary: An influence operation may prepare assets impersonating existing entities (both organisations and people) to further conceal its network identity and add a layer of legitimacy to its operation content. Existing entities may include authentic news outlets, public figures, organisations, or state entities. Users will more likely believe and less likely fact-check news from recognisable sources rather than unknown sites. An influence operation may use a wide variety of cyber techniques to impersonate a legitimate entity's website or social media account. This Technique was previously called Prepare Assets Impersonating Legitimate Entities

Funded by the European Union

None	**TA16: Establish Legitimacy** **T0099: Impersonate Existing Entity** **T0099.003: Impersonate Existing Organisation** Summary: A situation where a threat actor styles their online assets or content to mimic an existing organisation. This can be done to take advantage of peoples' trust in the organisation to increase narrative believability, to smear the organisation, or to make the organisation less trustworthy.
None	**TA16: Establish Legitimacy** **T0099: Impersonate Existing Entity** **T0099.004: Impersonate Existing Media Outlet** Summary: A situation where a threat actor styles their online assets or content to mimic an existing media outlet. This can be done to take advantage of peoples' trust in the outlet to increase narrative believability, to smear the outlet, or to make the outlet less trustworthy.
None	**TA16: Establish Legitimacy** **T0099: Impersonate Existing Entity** **T0099.005: Impersonate Existing Official** Summary: A situation where a threat actor styles their online assets or content to impersonate an official (including government officials, organisation officials, etc).
None	**TA16: Establish Legitimacy** **T0099: Impersonate Existing Entity** **T0099.006: Impersonate Existing Influencer** Summary: A situation where a threat actor styles their online assets or content to impersonate an influencer or celebrity, typically to exploit users' existing faith in the impersonated target.

As part of this change Astroturfing was renamed, and transitioned from a Subtechnique of Prepare Assets Impersonating Legitimate Entities to being its own top-level Technique.

Previous	Updated

Funded by the European Union

TA16: Establish Legitimacy **T0099: Prepare Assets Impersonating Legitimate Entities** **T0099.001: Astroturfing** Summary: Astroturfing occurs when an influence operation disguises itself as grassroots movement or organization that supports operation narratives. Unlike butterfly attacks, astroturfing aims to increase the appearance of popular support for the operation cause and does not infiltrate existing groups to discredit their objectives	**TA16: Establish Legitimacy** **T0142: Fabricate Grassroots Movement** Summary: This technique, sometimes known as "astroturfing", occurs when an influence operation disguises itself as a grassroots movement or organisation that supports operation narratives. Astroturfing aims to increase the appearance of popular support for an evolving grassroots movement in contrast to "Utilise Butterfly Attacka", which aims to discredit an existing grassroots movement. This Technique was previously called Astroturfing, and used the ID T0099.001

TA08: Conduct Pump Priming

Moved T0113: Employ Commercial Analytics Firms
Moved T0039: Bait Legitimate Influencers

We've heard feedback from a lot of our users that *Conduct Pump Priming* is a confusing Tactic which doesn't provide much value. Based on this, we're beginning to update Techniques housed there with a view to retiring *Conduct Pump Priming* entirely in future updates.

Employ Commercial Analytic Firms previously sat under *Conduct Pump Priming* in the *Execute* Phase, but deep analysis of a target audience is likely something that would be undertaken much earlier in an operation.

We considered both *Target Audience Analysis*, *Microtarget, and Establish Assets* as new homes for the Technique. We landed on the latter given that the firm is an asset that threat actors are establishing by employing them.

Previous	Updated
P03: Execute **TA08: Conduct Pump Priming** **T0113: Employ Commercial Analytics Firms** Summary: Commercial analytic firms collect data on target audience activities and evaluate the data to detect trends, such as content receiving high click-rates. An influence operation may employ commercial analytic firms to facilitate external collection on its target audience, complicating attribution efforts and better tailoring the content to audience preferences.	**P01: Plan** **TA15: Establish Assets** **T0113: Employ Commercial Analytics Firms** Summary: Commercial analytic firms collect data on target audience activities and evaluate the data to detect trends, such as content receiving high click-rates. An influence operation may employ commercial analytic firms to facilitate external collection on its target audience, complicating attribution efforts and better tailoring the content to audience preferences.

Bait Legitimate Influencers describes trying to trick existing influencers into amplifying campaign content to their network, but this doesn't match its parent Tactic *Conduct Pump Priming.* We considered several new Tactics for the Technique, including *Deliver Content* (it's a method of delivering content) and *Microtarget* (it targets very specific individuals), but we landed on *Maximise Exposure*, as it most closely matches the Technique's goal of exposing campaign content to a wider audience.

Funded by the European Union

We also took the opportunity to refine the Tactic's Name and Summary, for conciseness and clarity.

Previous	Updated
P03: Execute **TA08: Conduct Pump Priming** **T0039: Bait Legitimate Influencers** Summary: Credibility in a social media environment is often a function of the size of a user's network. "Influencers" are so-called because of their reach, typically understood as: 1) the size of their network (i.e. the number of followers, perhaps weighted by their own influence); and 2) The rate at which their comments are re-circulated (these two metrics are related). Add traditional media players at all levels of credibility and professionalism to this, and the number of potential influencial carriers available for unwitting amplification becomes substantial. By targeting high-influence people and organisations in all types of media with narratives and content engineered to appeal their emotional or ideological drivers, influence campaigns are able to add perceived credibility to their messaging via saturation and adoption by trusted agents such as celebrities, journalists and local leaders.	**P03: Execute** **TA17: Maximise Exposure** **T0039: Bait Influencer** Summary: Influencers are people on social media platforms who have large audiences. Threat Actors can try to trick Influencers such as celebrities, journalists, or local leaders who aren't associated with their campaign into amplifying campaign content. This gives them access to the Influencer's audience without having to go through the effort of building it themselves, and it helps legitimise their message by associating it with the Influencer, benefitting from their audience's trust in them.

Small Changes

Updated T0097.001: Backstop Personas
Updated T0104.002: Dating Apps
Updated T0049: Flooding the Information Space
Updated TA15: Establish Social Assets
Updated TA05: Microtarget

T0097.001 Backstop Personas has been renamed to *T0097.001 Produce Evidence for Persona* in an effort to reduce reliance on industry terminology and make Techniques clearer at the framework level. It has also received an updated summary.

Previous	Updated

Funded by the European Union

TA16: Establish Legitimacy **T0097: Create Persona** **T0097.001: Backstop Personas** Summary: Create other assets/dossier/cover/fake relationships and/or connections or documents, sites, bylines, attributions, to establish/augment/inflate crediblity/believability	**TA16: Establish Legitimacy** **T0097: Create Persona** **T0097.001: Produce Evidence for Persona** Summary: People may produce evidence which supports the persona they are deploying (T0097) (aka "backstopping" the persona). This Technique covers situations where evidence is developed or produced as part of an influence operation to increase the perceived legitimacy of a persona used during IO, including creating accounts for the same persona on multiple platforms. The use of personas (T0097), and providing evidence to improve people's perception of one's persona (T0097.001), are not necessarily malicious or inauthentic. However, sometimes people use personas to increase the perceived legitimacy of narratives for malicious purposes. This Technique was previously called Backstop Personas.

We've added a summary to *T0104.002: Dating App*, and introduced a new Incident showing the use of dating apps in an operation.

Previous	Updated
TA07: Select Channels and Affordances **T0104: Social Networks** **T0104.002: Dating Apps** Summary: *None*	**TA07: Select Channels and Affordances** **T0104: Social Networks** **T0104.002: Dating App** Summary: "Dating App" refers to any platform (or platform feature) in which the ostensive purpose is for users to develop a physical/romantic relationship with other users. Threat Actors can exploit users' quest for love to trick them into doing things like revealing sensitive information or giving them money. Examples include Tinder, Bumble, Grindr, Facebook Dating, Tantan, Badoo, Plenty of Fish, hinge, LOVOO, OkCupid, happn, and Mamba.

T0049: Flooding the Information Space's definition has been tweaked to allow for users to tag the flooding of information spaces other than social media feeds, and has had its name shortened slightly.

Previous	Updated

Funded by the European Union

TA17: Maximise Exposure **T0049: Flooding the Information Space** Summary: Flooding and/or mobbing social media channels feeds and/or hashtag with excessive volume of content to control/shape online conversations and/or drown out opposing points of view. Bots and/or patriotic trolls are effective tools to acheive this effect	**TA17: Maximise Exposure** **T0049: Flood Information Space** Summary: Flooding sources of information (e.g. Social Media feeds) with a high volume of inauthentic content. This can be done to control/shape online conversations, drown out opposing points of view, or make it harder to find legitimate information. Bots and/or patriotic trolls are effective tools to achieve this effect

TA15 has been renamed from *Establish Social Assets* to *Establish Assets,* to match the wide variety of asset types the Tactic encapsulates. Its summary has not been changed at this time.

Previous	Updated
TA15: Establish Social Assets Summary: Establishing information assets generates messaging tools, including social media accounts, operation personnel, and organisations, including directly and indirectly managed assets. For assets under their direct control, the operation can add, change, or remove these assets at will. Establishing information assets allows an influence operation to promote messaging directly to the target audience without navigating through external entities. Many online influence operations create or compromise social media accounts as a primary vector of information dissemination.	**TA15: Establish Assets** Summary: Establishing information assets generates messaging tools, including social media accounts, operation personnel, and organisations, including directly and indirectly managed assets. For assets under their direct control, the operation can add, change, or remove these assets at will. Establishing information assets allows an influence operation to promote messaging directly to the target audience without navigating through external entities. Many online influence operations create or compromise social media accounts as a primary vector of information dissemination.

TA05: Microtarget's summary has been updated to better differentiate it from *TA13: Target Audience Analysis*.

Previous	Updated
TA05: Microtarget: Summary: Target very specific populations of people	**TA05: Microtarget:** Summary: Actions taken which help target content to specific audiences identified and analysed as part of *TA13: Target Audience Analysis*

Incidents

Incidents in the DISARM Red Framework are intended to provide examples of real-world use of Techniques, to help users better understand and contextualise behaviours. This update sees three new Incidents introduced covering the following Techniques:

- T0141.001: Acquire Compromised Account
- T0141.002: Acquire Compromised Website
- T0104.002: Dating App

Funded by the European Union

Request for Feedback

TA07: Select Channels and Affordances

Issue: Overlap in platforms between T0103: Livestream, T0104: Social Networks, and T0104: Media Sharing Networks

We're aware that there is some overlap between Techniques within *TA07: Select Channels and Affordances,* particularly in *T0104* and *T0105*. For example, a platform like Instagram might fit in *T0104.001: Mainstream Social Networks,* but also could be categorised under *T0105: Media Sharing Networks,* any one of its sub-techniques for *Photo, Video,* and *Audio sharing,* and even *T0103: Livestream* and its sub-techniques.

DISARM is working on getting its framework integrated with STIX and OpenCTI. As part of this work, users will be able to tag the specific platform that an operation is using by choosing from a STIX Open Vocabulary. As such we need to think about what the purpose of *TA07: Select Channels and Affordances* is in a post-STIX DISARM. One avenue we're considering is listing platform features exploited by threat actors, rather than platform grouping.

There are examples of both in the framework under *T0104; T0104.001: Mainstream Social Networks* is a 'platform grouping' style technique, listing platforms considered 'mainstream', where *T0104.002: Dating Apps* is a 'platform feature' style technique, describing a feature which could exist on many platforms.

Feature focused techniques are less likely to become outdated, and provide a useful aggregation which can exist alongside STIX's individual platform tagging capabilities; we could use STIX tag that Facebook was used, but indicate that the "Facebook Dating" feature specifically was exploited using *T0104.002,* or "Facebook Live" using *T0103.001: Video Livestream,* etc.

This change would require a large rework of many Techniques and sub-techniques in TA07. Before undertaking such an effort, we wanted to give the community a chance to tell us what version of TA07 would most useful;

- **What kinds of Techniques do you think would be most useful under TA07?**
- **What kind of platform features would you like to catalogue using DISARM?**
- **Do you have any alternative suggestions for how we should restructure TA07?**

Please reach out on info@disarm.foundation if you have any feedback, or would be interested in having a discussion with a member of the DISARM team.

Funded by the European Union

Annex 2 - DISARM Version 1.5 Release Notes

DISARM Red Version 1.5 - Patch Notes

In this update we begin our work to make DISARM's Red framework interoperable with Meta's Online Operations Kill Chain (MOOKC), as part of DISARM's work in the ADAC.IO project to improve how the counter influence community collaborates in the fight against FIMI (Foreign Information Manipulation and Interference). This update focuses on mapping the "Disguising Assets'' section of MOOKC to the DISARM Red Framework.

We will first provide an overview of MOOKC's Disguising Assets, then discuss issues which arose when attempting 1-1 mapping, moving on to introducing the new Techniques being added to DISARM's Red framework in this update.

Contents:

Funded by the European Union

T0009.001: Utilise Academic/Pseudoscientific Justifications
T0142: Fabricate Grassroots Movement

Reworked

T0097: Create Personas
T0097.001: Produce Evidence for Persona
T0099: Impersonate Existing Entity
T0099.002: Spoof/Parody Account/Site

Small Changes

T0085.001: Develop AI-Generated Text
T0086.002: Develop AI-Generated Images
T0104.002: Dating App

Meta's Disguising Assets Overview

The following is a list of all items in MOOKC section 1: Disguising Assets. Items highlighted in red have not been transferred to DISARM Red at this time.

1. Disguising assets
    - 1.1. Adopting visual disguise
        - 1.1.1. Copying profile pictures
        - 1.1.2. Using profile pictures created using generative adversarial networks (GAN)
        - 1.1.3. Adopting Visual Brand
        - 1.1.4. Using animals as profile picture
        - 1.1.5. Using scenery as profile picture
        - 1.1.6. Using cartoon as profile picture
    - 1.2. Posing as non-existent person
        - 1.2.1. Posing as person seeking romance
        - 1.2.2. Posing as fictional journalist
        - 1.2.3. Posing as fictional activist
        - 1.2.4. Posing as fictional hacktivist
        - 1.2.5. Posing as fictional military personnel
        - 1.2.6. Posing as fictional recruiter / potential employer
        - 1.2.7. Creating fictitious byline
        - 1.2.8. Posing as fictional person in target region
    - 1.3. Posing as non-existent institution
        - 1.3.1. Creating fictitious news outlet
        - 1.3.2. Creating fictitious NGO
    - 1.4. Impersonating real person
        - 1.4.1. Impersonating researcher or think tanker
        - 1.4.2. Using duplicate accounts
    - 1.5. Impersonating real institution
        - 1.5.1. Impersonating news website
        - 1.5.2. Impersonating government institution
        - 1.5.3. Impersonating think tank

Funded by the European Union

1.5.4. Impersonating commercial company
1.6. Disguising malware sites [x5]
1.7. Disguising malicious apps [x3]
1.8. Backstopping
1.8.1. Backstopping fictitious individual across multiple websites
1.8.2. Backstopping fictitious brand or organisation across multiple websites

Issues preventing 1-1 Meta Mapping

In the examples provided above, Meta analysts assert that assets are being disguised, and that accounts are presenting as either fictitious people (e.g. "pose as fictional journalist" or "create fictitious news outlet") or as impersonations. Analysts can assert which type of fictitious or impersonated persona is being presented.

Feedback provided to DISARM in a recent survey of the counter-FIMI community suggested that many analysts avoided using existing DISARM Techniques which asserted things which didn't necessarily align with what they were seeing. For example, some analysts avoid applying the Technique "Create Inauthentic Accounts", because they cannot justify labelling accounts as 'inauthentic' with the evidence available to them.

Meta has access to detailed technical indicators which will help them confidently assert whether an account is presenting a fictitious or impersonated persona. However, DISARM needs to support analysts who cannot determine the legitimacy of a given persona when they first encounter it. This meant separating out the assertion that an account was presenting a "fictitious" or "impersonated' persona, so that analysts have the ability to document how an account is presenting.

For example, analysts can use DISARM Red's new *Journalist Persona (T0097.102)* sub-technique to assert that an account is presenting as a journalist, and can then move on to investigating whether the persona has been fabricated, or an impersonation (using the new *T0143.002: Fabricated Persona* or *T0143.004: Impersonated Persona*). This approach also allows for documenting the use of peoples' *Authentic Persona (T0143.001)* (e.g. a democratically elected politician amplifying a false narrative online), and of parody (*T0143.003: Parody Persona*).

DISARM approach

We touched upon some new sub-techniques in the last paragraph. In this section we'll introduce the new top level techniques

T0097: Present Persona

T0097: Present Persona allows analysts to document the type of persona presented without asserting the persona's legitimacy. Analysts will first identify how the account is presenting itself (e.g. "Journalist Persona"), and then they can investigate the persona's validity using *T0143: Persona Legitimacy*. Identifying the persona is (in our experience) a much quicker task than figuring out if they're real or not, so separating out the two should be useful.

An added benefit of this approach is that by separating the assertion of validity from the type of persona, we make it so there's never a situation where there's a persona type under "fabricated person" but not under "impersonated

Funded by the European Union

person", and reduces the total number of Techniques we need to add (when we add a new persona, we just add "that persona", instead of "fabricate that persona", "impersonate that persona", and "authentic version of that persona"). Overall this approach should mean fewer Techniques in the framework, which people have asked for, and gives us the opportunity to easily expand the roster of personas based on user feedback.

Link to detailed changes

T0143: Persona Legitimacy

T0143: Persona Legitimacy allows analysts to document the legitimacy of the persona; whether it's authentic, fabricated, an impersonation, or a parody. Looking back to MOOKC, item *1.2.2: Posing as a fictional* journalist can be documented using a combination of *T0097.102: Journalist Persona* and *T0143.002: Fabricated Persona*.

Link to detailed changes

T0144: Persona Legitimacy Evidence

T0097: Presdented Personas is a rework of the existing technique T0097: Create Personas. As part of this rework, the sub-technique *T0097.001: Produce Evidence for Persona* has been reworked into the technique *T0144: Persona Legitimacy Evidence*. This technique contains methods used for "backstopping" the Persona, such as presenting the same persona across different platforms, or using a template to quickly fabricate many personas.

Link to detailed changes

T0145: Establish Account Imagery

T0145: Establish Account Imagery covers the behaviours described in MOOKC's *Adopting Visual Disguise,* but enables analysts to describe accounts' imagery without asserting that it's a disguise.

The MOOKC specifies the use of profile pictures (e.g. *Using animals as profile picture)*, where the new DISARM Red techniques allow for other types of account imagery such as profile headers (e.g. T0145.003: Animal Account Imagery).

Link to detailed changes

New DISARM Documentation

In addition to these patch notes, new documentation has been created, which will be updated with each iteration on the DISARM Red framework.

Funded by the European Union

Mapping DISARM to Meta

The initial approach for helping analysts map DISARM techniques to Meta was to explain mapping within the descriptions of relevant techniques and sub-techniques. This added too much complication to the techniques' descriptions, and given users' feedback that the framework needs to be more approachable, the decision was made to cut this from the descriptions.

Instead, DISARM has produced documentation which describes how MOOKC framework items can be mapped to the DISARM Red framework.

Deprecated and Changed Techniques and Sub-techniques

As DISARM introduces more changes to the framework, it's important to help analysts keep track of changes which have been made, particularly where existing techniques are updated or removed.

This documentation collates all changes within framework patch notes which cover any kind of update to existing DISARM Red framework items.

Disarm Changes Overview

Items below are colour coded to help indicate which items have been added, removed, and reworked, with other framework items listed helping users locate the changes within the existing Red framework.

New Additions

- TA16: Establish Legitimacy
    - T0097: Present Persona
        - T0097.100: Individual Persona
            - T0097.101: Local Persona
            - T0097.102: Journalist Persona
            - T0097.103: Activist Persona
            - T0097.104: Hacktivist Persona
            - T0097.105: Military Personnel Persona
            - T0097.106: Recruiter Persona
            - T0097.107: Researcher Persona
            - T0097.108: Expert Persona
            - T0097.109: Romantic Suitor Persona
            - T0097.110: Party Official Persona
            - T0097.111: Government Official Persona
            - T0097.112: Government Employee Persona
        - T0097.200: Institutional Persona
            - T0097.201: Local Institution Persona
            - T0097.202: News Outlet Persona

Funded by the European Union

                - T0097.203: Fact Checking Organisation Persona
                - T0097.204: Think Tank Persona
                - T0097.205: Business Persona
                - T0097.206: Government Institution Persona
                - T0097.207: NGO Persona
                - T0097.208: Social Cause Persona
        - T0143: Persona Legitimacy
            - T0143.001: Authentic Persona
            - T0143.002: Fabricated Persona
            - T0143.003: Impersonated Persona
            - T0143.004: Parody Persona
        - T0144: Persona Legitimacy Evidence
            - T0144.001: Persona Presented across Platforms
            - T0144.002: Persona Template
- TA15: Establish Assets
    - T0145: Establish Account Imagery
        - T0145.001: Copy Account Imagery
        - T0145.002: AI-Generated Account Imagery
        - T0145.003: Animal Account Imagery
        - T0145.004: Scenery Account Imagery
        - T0145.005: Illustrated Character Account Imagery
        - T0145.006: Stock Image Account Imagery
        - T0145.007: Attractive Person Account Imagery

New Non-Meta Techniques

- TA06: Develop Content
    - T0085: Develop Text-Based Content
        - T0085.008: Machine Translated Text

Removed Techniques and Sub-Techniques

- TA16: Establish Legitimacy
    - T0009: Create Fake Experts
        - T0009.001: Utilise Academic/Pseudoscientific Justifications
    - T0099: Impersonate Existing Entity
        - T0099.002: Spoof/Parody Account/Site
        - T0099.003: Impersonate Existing Organisation
        - T0099.004: Impersonate Existing Official
        - T0099.005: Impersonate Existing Media Outlet
        - T0099.006: Impersonate Existing Influencer
    - T0142: Fabricate Grassroots Movement

Reworked Techniques and Sub-Techniques

- T0097: Create Personas

Funded by the European Union

    - T0097.001: Present Evidence for Persona
- T0099: Impersonate Existing Entity
    - T0099.002: Spoof/Parody Account/Site

Small Changes

- T0085: Develop Text-Based Content
    - T0085.001: Develop AI-Generated Text
- T0086: Develop Image-Based Content
    - T0086.002: Develop AI-Generated Images
- T0104: Social Networks
    - T0104.002: Dating App

New Incidents

We've heard users' feedback that they would like more real-world examples of techniques being used in the wild, to help them put theory into context. In this update 31 new incidents are being introduced to the Red framework, providing 128 real-world examples of techniques' usage. These new incidents predominantly focus on supporting users in understanding the newly introduced framework items, but some examples of existing techniques have also been added.

Detailed Disarm Changes

New Additions

T0097: Present Persona

T0097: Present Persona allows analysts to document the type of persona presented without asserting the persona's legitimacy.

Link to technique introduction

New DISARM Techniques and Sub-Techniques
TA16: Establish Legitimacy **T0097: Present Persona** Summary: This Technique contains different types of personas commonly taken on by threat actors during influence operations. Analysts should use T0097's sub-techniques to document the type of persona which an account is presenting. For example, an account which describes itself as being a journalist can be tagged with T0097.102: Journalist Persona. Personas presented by individuals include:

Funded by the European Union

T0097.100: Individual Persona
T0097.101: Local Persona
T0097.102: Journalist Persona
T0097.103: Activist Persona
T0097.104: Hacktivist Persona
T0097.105: Military Personnel Persona
T0097.106: Recruiter Persona
T0097.107: Researcher Persona
T0097.108: Expert Persona
T0097.109: Romantic Suitor Persona
T0097.110: Party Official Persona
T0097.111: Government Official Persona
T0097.112: Government Employee Persona

This Technique also houses institutional personas commonly taken on by threat actors:

T0097.200: Institutional Persona
T0097.201: Local Institution Persona
T0097.202: News Outlet Persona
T0097.203: Fact Checking Organisation Persona
T0097.204: Think Tank Persona
T0097.205: Business Persona
T0097.206: Government Institution Persona
T0097.207: NGO Persona
T0097.208: Social Cause Persona

By using a persona, a threat actor is adding the perceived legitimacy of the persona to their narratives and activities.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.100: Individual Persona
Summary: This sub-technique can be used to indicate that an entity is presenting itself as an individual. If the person is presenting themselves as having one of the personas listed below then these sub-techniques should be used instead, as they indicate both the type of persona they presented and that the entity presented itself as an individual:

T0097.101: Local Persona
T0097.102: Journalist Persona
T0097.103: Activist Persona
T0097.104: Hacktivist Persona
T0097.105: Military Personnel Persona
T0097.106: Recruiter Persona
T0097.107: Researcher Persona
T0097.108: Expert Persona
T0097.109: Romantic Suitor Persona
T0097.110: Party Official Persona
T0097.111: Government Official Persona
T0097.112: Government Employee Persona

Funded by the European Union

TA16: Establish Legitimacy
T0097: Present Persona
T0097.101: Local Persona
Summary: A person with a local persona presents themselves as living in a particular geography or having local knowledge relevant to a narrative.

While presenting as a local is not an indication of inauthentic behaviour, an influence operation may have its narratives amplified by people presenting as local to a target area. Threat actors can fabricate locals (T0143.002: Fabricated Persona, T0097.101: Local Persona) to add credibility to their narratives, or to misrepresent the real opinions of locals in the area.

People who are legitimate locals (T0143.001: Authentic Persona, T0097.101: Local Persona) can use their persona for malicious purposes, or be exploited by threat actors. For example, someone could take money for using their position as a local to provide legitimacy to a false narrative or be tricked into doing so without their knowledge.

Associated Techniques and Sub-techniques
T0097.201: Local Institution Persona: Analysts should use this sub-technique to catalogue cases where an institution is presenting as a local, such as a local news organisation or local business.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.102: Journalist Persona
Summary: A person with a journalist persona presents themselves as a reporter or journalist delivering news, conducting interviews, investigations etc.

While presenting as a journalist is not an indication of inauthentic behaviour, an influence operation may have its narratives amplified by people presenting as journalists. Threat actors can fabricate journalists to give the appearance of legitimacy, justifying the actor's requests for interviews, etc (T0143.002: Fabricated Persona, T0097.102: Journalist Persona).

People who have legitimately developed a persona as a journalist (T0143.001: Authentic Persona, T0097.102: Journalist Persona) can use it for malicious purposes, or be exploited by threat actors. For example, someone could take money for using their position as a trusted journalist to provide legitimacy to a false narrative or be tricked into doing so without the journalist's knowledge.

Associated Techniques and Sub-techniques
T0097.202: News Organisation Persona: People with a journalist persona may present as being part of a news organisation.
T0097.101: Local Persona: People with a journalist persona may present themselves as local reporters.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.103: Activist Persona
Summary: A person with an activist persona presents themselves as an activist; an individual who campaigns for a political cause, organises related events, etc.

While presenting as an activist is not an indication of inauthentic behaviour, an influence operation may have its narratives amplified by people presenting as activists. Threat actors can fabricate activists to give the appearance of popular support for an evolving grassroots movement (see T0143.002: Fabricated Persona, T0097.103: Activist Persona).

Funded by the European Union

People who are legitimate activists can use this persona for malicious purposes, or be exploited by threat actors. For example, someone could take money for using their position as an activist to provide visibility to a false narrative or be tricked into doing so without their knowledge (T0143.001: Authentic Persona, T0097.103: Activist Persona).

Associated Techniques and Sub-techniques
T0097.104: Hacktivist Persona: Analysts should use this sub-technique to catalogue cases where an individual is presenting themselves as someone engaged in activism who uses technical tools and methods, including building technical infrastructure and conducting offensive cyber operations, to achieve their goals.
T0097.207: NGO Persona: People with an activist persona may present as being part of an NGO.
T0097.208: Social Cause Persona: Analysts should use this sub-technique to catalogue cases where an online account is presenting as posting content related to a particular social cause, while not presenting as an individual.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.104: Hacktivist Persona
Summary: A person with a hacktivist persona presents themselves as an activist who conducts offensive cyber operations or builds technical infrastructure for political purposes, rather than the financial motivations commonly attributed to hackers; hacktivists are hacker activists who use their technical knowledge to take political action.

Hacktivists can build technical infrastructure to support other activists, including secure communication channels and surveillance and censorship circumvention. They can also conduct DDOS attacks and other offensive cyber operations, aiming to take down digital assets or gain access to proprietary information. An influence operation may use hacktivist personas to support their operational narratives and legitimise their operational activities.

Fabricated Hacktivists are sometimes referred to as “Faketivists”.

Associated Techniques and Sub-techniques
T0097.103: Activist Persona: Analysts should use this sub-technique to catalogue cases where an individual is presenting themselves as someone engaged in activism but doesn’t present themselves as using technical tools and methods to achieve their goals.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.105: Military Personnel Persona
Summary: A person with a military personnel persona presents themselves as a serving member or veteran of a military organisation operating in an official capacity on behalf of a government.

While presenting as military personnel is not an indication of inauthentic behaviour, an influence operation may have its narratives amplified by people presenting as military personnel. Threat actors can fabricate military personnel (T0143.002: Fabricated Persona, T0097.105: Military Personnel Persona) to pose as experts on military topics, or to discredit geopolitical adversaries by pretending to be one of their military personnel and spreading discontent.

People who have legitimately developed a military persona (T0143.001: Authentic Persona, T0097.105: Military Personnel Persona) can use it for malicious purposes, or be exploited by threat actors. For example, someone could take money for using their position as a member of the military to provide legitimacy to a false narrative or be tricked into doing so without their knowledge.

Funded by the European Union

TA16: Establish Legitimacy
T0097: Present Persona
T0097.106: Recruiter Persona
Summary:A person with a recruiter persona presents themselves as a potential employer or provider of freelance work.

While presenting as a recruiter is not an indication of inauthentic behaviour, threat actors fabricate recruiters (T0143.002: Fabricated Persona, T0097.106: Recruiter Persona) to justify asking for personal information from their targets or to trick targets into working for the threat actors (without revealing who they are).

Associated Techniques and Sub-techniques
T0097.205: Business Persona: People with a recruiter persona may present as being part of a business which they are recruiting for.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.107: Researcher Persona
Summary: A person with a researcher persona presents themselves as conducting research (e.g. for academic institutions, or think tanks), or having previously conducted research.

While presenting as a researcher is not an indication of inauthentic behaviour, an influence operation may have its narratives amplified by people presenting as researchers. Threat actors can fabricate researchers (T0143.002: Fabricated Persona, T0097.107: Researcher Persona) to add credibility to their narratives.

People who are legitimate researchers (T0143.001: Authentic Persona, T0097.107: Researcher Persona) can use their persona for malicious purposes, or be exploited by threat actors. For example, someone could take money for using their position as a Researcher to provide legitimacy to a false narrative or be tricked into doing so without their knowledge.

Associated Techniques and Sub-techniques
T0097.204: Think Tank Persona: People with a researcher persona may present as being part of a think tank.
T0097.108: Expert Persona: People who present as researching a given topic are likely to also present as having expertise in the area.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.108: Expert Persona
Summary: A person with an expert persona presents themselves as having expertise or experience in a field. Commonly the persona's expertise will be called upon to add credibility to a given narrative.

While presenting as an expert is not an indication of inauthentic behaviour, an influence operation may have its narratives amplified by people presenting as experts. Threat actors can fabricate experts (T0143.002: Fabricated Persona, T0097.107: Researcher Persona) to add credibility to their narratives.

People who are legitimate experts (T0143.001: Authentic Persona, T0097.107: Researcher Persona) can make mistakes, use their persona for malicious purposes, or be exploited by threat actors. For example, someone could take money for using their position as an expert to provide legitimacy to a false narrative or be tricked into doing so without their knowledge.

Associated Techniques and Sub-techniques
T0097.107: Researcher Persona: People who present as experts may also present as conducting or having

Funded by the European Union

conducted research into their specialist subject.
T0097.204: Think Tank Persona: People with an expert persona may present as being part of a think tank.

TA16: Establish Legitimacy
T0097: Present Persona
T0097:109 Romantic Suitor Persona
Summary: A person with a romantic suitor persona presents themselves as seeking a romantic or physical connection with another person.

While presenting as seeking a romantic or physical connection is not an indication of inauthentic behaviour, threat actors can use dating apps, social media channels or dating websites to fabricate romantic suitors to lure targets they can blackmail, extract information from, deceive or trick into giving them money (T0143.002: Fabricated Persona, T0097.109: Romantic Suitor Persona).

Honeypotting in espionage and Big Butchering in scamming are commonly associated with romantic suitor personas.

Associated Techniques and Sub-techniques
T0104.002: Dating App: Analysts can use this sub-technique for tagging cases where an account has been identified as using a dating platform.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.110: Party Official Persona
Summary: A person who presents as an official member of a political party, such as leaders of political parties, candidates standing to represent constituents, and campaign staff.

Presenting as an official of a political party is not an indication of inauthentic behaviour, however threat actors may fabricate individuals who work in political parties to add credibility to their narratives (T0143.002: Fabricated Persona, T0097.110: Party Official Persona). They may also impersonate existing officials of political parties (T0143.003: Impersonated Persona, T0097.110: Party Official Persona).

Legitimate members of political parties could use their persona for malicious purposes, or be exploited by threat actors (T0143.001: Authentic Persona, T0097.110: Party Official Persona). For example, an electoral candidate could take money for using their position to provide legitimacy to a false narrative, or be tricked into doing so without their knowledge.

Associated Techniques and Sub-techniques
T0097.111: Government Official Persona: Analysts should use this sub-technique to catalogue cases where an individual is presenting as a member of a government.

Some party officials will also be government officials. For example, in the United Kingdom the head of government is commonly also the head of their political party.

Some party officials won't be government officials. For example, members of a party standing in an election, or party officials who work outside of government (e.g. campaign staff).

TA16: Establish Legitimacy
T0097: Present Persona
T0097.111: Government Official Persona
Summary: A person who presents as an active or previous government official has the government official

Funded by the European Union

persona. These are officials serving in government, such as heads of government departments, leaders of countries, and members of government selected to represent constituents.

Presenting as a government official is not an indication of inauthentic behaviour, however threat actors may fabricate individuals who work in government to add credibility to their narratives (T0143.002: Fabricated Persona, T0097.111: Government Official Persona). They may also impersonate existing members of government (T0143.003: Impersonated Persona, T0097.111: Government Official Persona).

Legitimate government officials could use their persona for malicious purposes, or be exploited by threat actors (T0143.001: Authentic Persona, T0097.111: Government Official Persona). For example, a government official could take money for using their position to provide legitimacy to a false narrative, or be tricked into doing so without their knowledge.

Associated Techniques and Sub-techniques
T0097.110: Party Official Persona: Analysts should use this sub-technique to catalogue cases where an individual is presenting as a member of a political party.

Not all government officials are political party officials (such as outside experts brought into government) and not all political party officials are government officials (such as people standing for office who are not yet working in government).

T0097.206: Government Institution Persona: People presenting as members of a government may also represent a government institution which they are associated with.

T0097.112: Government Employee Persona: Analysts should use this sub-technique to document people presenting as professionals hired to serve in government institutions and departments, not officials selected to represent constituents, or assigned official roles in government (such as heads of departments).

TA16: Establish Legitimacy
T0097: Present Persona
T0097.112: Government Employee Persona
Summary: A person who presents as an active or previous civil servant has the government employee persona. These are professionals hired to serve in government institutions and departments, not officials selected to represent constituents, or assigned official roles in government (such as heads of departments).

Presenting as a government employee is not an indication of inauthentic behaviour, however threat actors may fabricate individuals who work in government to add credibility to their narratives (T0143.002: Fabricated Persona, T0097.112: Government Employee Persona). They may also impersonate existing government employees (T0143.003: Impersonated Persona, T0097.112: Government Employee Persona).

Legitimate government employees could use their persona for malicious purposes, or be exploited by threat actors (T0143.001: Authentic Persona, T0097.112: Government Employee Persona). For example, a government employee could take money for using their position to provide legitimacy to a false narrative, or be tricked into doing so without their knowledge.

Associated Techniques and Sub-techniques
T0097.111: Government Official Persona: Analysts should use this technique to document people who present as an active or previous government official, such as heads of government departments, leaders of countries, and members of government selected to represent constituents.

T0097.206: Government Institution Persona: People presenting as members of a government may also present a

Funded by the European Union

government institution which they are associated with.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.200: Institutional Persona
Summary: This Technique can be used to indicate that an entity is presenting itself as an institution. If the organisation is presenting itself as having one of the personas listed below then these Techniques should be used instead, as they indicate both that the entity presented itself as an institution, and the type of persona they presented:

T0097.201: Local Institution Persona
T0097.202: News Outlet Persona
T0097.203: Fact Checking Organisation Persona
T0097.204: Think Tank Persona
T0097.205: Business Persona
T0097.206: Government Institution Persona
T0097.207: NGO Persona
T0097.208: Social Cause Persona

TA16: Establish Legitimacy
T0097: Present Persona
T0097.201: Local Institution Persona
Summary: Institutions which present themselves as operating in a particular geography, or as having local knowledge relevant to a narrative, are presenting a local institution persona.

While presenting as a local institution is not an indication of inauthentic behaviour, threat actors may present themselves as such (T0143.002: Fabricated Persona, T0097.201: Local Institution Persona) to add credibility to their narratives, or misrepresent the real opinions of locals in the area.

Legitimate local institutions could use their persona for malicious purposes, or be exploited by threat actors (T0143.001: Authentic Persona, T0097.201: Local Institution Persona). For example, a local institution could take money for using their position to provide legitimacy to a false narrative, or be tricked into doing so without their knowledge.

Associated Techniques and Sub-techniques
T0097.101: Local Persona: Institutions presenting as local may also present locals working within the organisation.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.202: News Outlet Persona
Summary: An institution with a news outlet persona presents itself as an organisation which delivers new information to its target audience.

While presenting as a news outlet is not an indication of inauthentic behaviour, an influence operation may have its narratives amplified by news organisations. Threat actors can fabricate news organisations (T0143.002: Fabricated Persona, T0097.202: News Outlet Persona), or they can impersonate existing news outlets (T0143.003: Impersonated Persona, T0097.202: News Outlet Persona).

Legitimate news organisations could use their persona for malicious purposes, or be exploited by threat actors (T0143.001: Authentic Persona, T0097.202: News Outlet Persona).

Funded by the European Union

Associated Techniques and Sub-techniques
T0097.102: Journalist Persona: Institutions presenting as news outlets may also present journalists working within the organisation.
T0097.201: Local Institution Persona: Institutions presenting as news outlets may present as being a local news outlet.
T0097.203: Fact Checking Organisation Persona: Institutions presenting as news outlets may also deliver a fact checking service (e.g. The UK's BBC News has the fact checking service BBC Verify). When an actor presents as the fact checking arm of a news outlet, they are presenting both a News Outlet Persona and a Fact Checking Organisation Persona.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.203: Fact Checking Organisation Persona
Summary: An institution with a fact checking organisation persona presents itself as an organisation which produces reports which assess the validity of others' reporting / statements.

While presenting as a fact checking organisation is not an indication of inauthentic behaviour, an influence operation may have its narratives amplified by fact checking organisations. Threat actors can fabricate fact checking organisations (T0143.002: Fabricated Persona, T0097.202: News Outlet Persona), or they can impersonate existing fact checking outlets (T0143.003: Impersonated Persona, T0097.202: News Outlet Persona).

Legitimate fact checking organisations could use their persona for malicious purposes, or be exploited by threat actors (T0143.001: Authentic Persona, T0097.202: News Outlet Persona).

Associated Techniques and Sub-techniques
T0097.102: Journalist Persona: Institutions presenting as fact checking organisations may also present journalists working within the organisation.
T0097.202: News Outlet Persona: Fact checking organisations may present as operating as part of a larger news outlet (e.g. The UK's BBC News has the fact checking service BBC Verify). When an actor presents as the fact checking arm of a news outlet, they are presenting both a News Outlet Persona and a Fact Checking Organisation Persona.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.204: Think Tank Persona
Summary: An institution with a think tank persona presents itself as a think tank; an organisation that aims to conduct original research and propose new policies or solutions, especially for social and scientific problems.

While presenting as a think tank is not an indication of inauthentic behaviour, think tank personas are commonly used by threat actors as a front for their operational activity (T0143.002: Fabricated Persona, T0097.204: Think Tank Persona). They may be created to give legitimacy to narratives and allow them to suggest politically beneficial solutions to societal issues.

Legitimate think tanks could have a political bias that they may not be transparent about, they could use their persona for malicious purposes, or they could be exploited by threat actors (T0143.001: Authentic Persona, T0097.204: Think Tank Persona). For example, a think tank could take money for using their position to provide legitimacy to a false narrative, or be tricked into doing so without their knowledge.

Associated Techniques and Sub-techniques

Funded by the European Union

T0097.107: Researcher Persona:</br> Institutions presenting as think tanks may also present researchers working within the organisation.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.205: Business Persona
Summary: An institution with a business persona presents itself as a for-profit organisation which provides goods or services for a price.

While presenting as a business is not an indication of inauthentic behaviour, business personas may be used by threat actors as a front for their operational activity (T0143.002: Fabricated Persona, T0097.205: Business Persona).

Threat actors may also impersonate existing businesses (T0143.003: Impersonated Persona, T0097.205: Business Persona) to exploit their brand or cause reputational damage.

Legitimate businesses could use their persona for malicious purposes, or be exploited by threat actors (T0143.001: Authentic Persona, T0097.205: Business Persona). For example, a business could take money for using their position to provide legitimacy to a false narrative, or be tricked into doing so without their knowledge.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.206: Government Institution Persona
Summary: Institutions which present themselves as governments, or government ministries, are presenting a government institution persona.

While presenting as a government institution is not an indication of inauthentic behaviour, threat actors may impersonate existing government institutions as part of their operation (T0143.003: Impersonated Persona, T0097.206: Government Institution Persona), to add legitimacy to their narratives, or discredit the government.

Legitimate government institutions could use their persona for malicious purposes, or be exploited by threat actors (T0143.001: Authentic Persona, T0097.206: Government Institution Persona). For example, a government institution could be used by elected officials to spread inauthentic narratives.

Associated Techniques and Sub-techniques
T0097.111: Government Official Persona: Institutions presenting as governments may also present officials working within the organisation.
T0097.112: Government Employee Persona: Institutions presenting as governments may also present employees working within the organisation.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.207: NGO Persona
Summary: Institutions which present themselves as an NGO (Non-Governmental Organisation), an organisation which provides services or advocates for public policy (while not being directly affiliated with any government), are presenting an NGO persona.

While presenting as an NGO is not an indication of inauthentic behaviour, NGO personas are commonly used by threat actors (such as intelligence services) as a front for their operational activity (T0143.002: Fabricated Persona, T0097.207: NGO Persona). They are created to give legitimacy to the influence operation and potentially infiltrate grassroots movements

Funded by the European Union

Legitimate NGOs could use their persona for malicious purposes, or be exploited by threat actors (T0143.001: Authentic Persona, T0097.207: NGO Persona). For example, an NGO could take money for using their position to provide legitimacy to a false narrative, or be tricked into doing so without their knowledge.

Associated Techniques and Sub-techniques:
T0097.103: Activist Persona: Institutions presenting as activist groups may also present activists working within the organisation.

TA16: Establish Legitimacy
T0097: Present Persona
T0097.208: Social Cause Persona
Summary: Online accounts which present themselves as focusing on a social cause are presenting the Social Cause Persona. Examples include accounts which post about current affairs, such as discrimination faced by minorities.

While presenting as an account invested in a social cause is not an indication of inauthentic behaviour, such personas have been used by threat actors to exploit peoples' legitimate emotional investment regarding social causes that matter to them (T0143.002: Fabricated Persona, T0097.208: Social Cause Persona).

Legitimate accounts focused on a social cause could use their persona for malicious purposes, or be exploited by threat actors (T0143.001: Authentic Persona, T0097.208: Social Cause Persona). For example, the account holders could take money for using their position to provide legitimacy to a false narrative, or be tricked into doing so without their knowledge.

Associated Techniques and Sub-techniques:
T0097.103: Activist Persona: Analysts should use this sub-technique to catalogue cases where an individual is presenting themselves as an activist related to a social cause. Accounts with social cause personas do not present themselves as individuals, but may have activists controlling the accounts.

T0143: Persona Legitimacy

T0143: Persona Legitimacy allows analysts to document the legitimacy of the persona; whether it's authentic, fabricated, an impersonation, or a parody.

Link to technique introduction

Proposed DISARM Additions
TA16: Establish Legitimacy **T0143: Persona Legitimacy** Summary: This Technique contains sub-techniques which analysts can use to assert whether an account is presenting an authentic, fabricated, or parody persona: T0143.001: Authentic Persona T0143.002: Fabricated Persona T0143.003: Impersonated Persona T0143.004: Parody Persona
TA16: Establish Legitimacy

Funded by the European Union

T0143: Persona Legitimacy
T0143.001: Authentic Persona
Summary: An individual or institution presenting a persona that legitimately matches who or what they are is presenting an authentic persona.

For example, an account which presents as being managed by a member of a country's military, and is legitimately managed by that person, would be presenting an authentic persona (T0143.001: Authentic Persona, T0097.105: Military Personnel).

Sometimes people can authentically present themselves as who they are while still participating in malicious/inauthentic activity; a legitimate journalist (T0143.001: Authentic Persona, T0097.102: Journalist Persona) may accept bribes to promote products, or they could be tricked by threat actors into sharing an operation's narrative.

TA16: Establish Legitimacy
T0143: Persona Legitimacy
T0143.002: Fabricated Persona
Summary: An individual or institution pretending to have a persona without any legitimate claim to that persona is presenting a fabricated persona, such as a person who presents themselves as a member of a country's military without having worked in any capacity with the military (T0143.002: Fabricated Persona, T0097.105: Military Personnel).

Sometimes real people can present entirely fabricated personas; they can use real names and photos on social media while also pretending to have credentials or traits they don't have in real life.

TA16: Establish Legitimacy
T0143: Persona Legitimacy
T0143.003: Impersonated Persona
Summary: Threat actors may impersonate existing individuals or institutions to conceal their network identity, add legitimacy to content, or harm the impersonated target's reputation. This Technique covers situations where an actor presents themselves as another existing individual or institution.

This Technique was previously called Prepare Assets Impersonating Legitimate Entities and used the ID T0099.

Associated Techniques and Sub-techniques
T0097: Present Persona: Analysts can use the sub-techniques of T0097: Present Persona to categorise the type of impersonation. For example, a document developed by a threat actor which falsely presented as a letter from a government department could be documented using T0085.004: Develop Document, T0143.003: Impersonated Persona, and T0097.206: Government Institution Persona.
T0145.001: Copy Account Imagery: Actors may take existing accounts' profile pictures as part of their impersonation efforts.

TA16: Establish Legitimacy
T0143: Persona Legitimacy
T0143:004: Parody Persona
Summary: Parody is a form of artistic expression that imitates the style or characteristics of a particular work, genre, or individual in a humorous or satirical way, often to comment on or critique the original work or subject matter. People may present as parodies to create humour or make a point by exaggerating or altering elements of the original, while still maintaining recognizable elements.

The use of parody is not an indication of inauthentic or malicious behaviour; parody allows people to present ideas or criticisms in a comedic or exaggerated manner, softening the impact of sensitive or contentious topics.

Funded by the European Union

Because parody is often protected as a form of free speech or artistic expression, it provides a legal and social framework for discussing controversial issues.

However, parody personas may be perceived as authentic personas, leading to people mistakenly believing that a parody account's statements represent the real opinions of a parodied target. Threat actors may also use the guise of parody to spread campaign content. Parody personas may disclaim that they are operating as a parody, however this is not always the case, and is not always given prominence.

Associated Techniques and Sub-techniques
T0097: Present Persona: Analysts can use the sub-techniques of T0097: Present Persona to categorise the type of parody. For example, an account presenting as a parody of a business could be documented using T0097.205: Business Persona and T0143.003: Parody Persona.
T0145.001: Copy Account Imagery: Actors may take existing accounts' profile pictures as part of their parody efforts.

T0144: Persona Legitimacy Evidence

T0144: Persona Legitimacy Evidence contains methods used to "backstopping" the Persona, such as presenting the same persona across different platforms, or using a template to quickly fabricate many personas.

Link to technique introduction

Proposed DISARM Additions
TA16: Establish Legitimacy **T0144: Persona Legitimacy Evidence** Summary: This Technique contains behaviours which might indicate whether a persona is legitimate, a fabrication, or a parody. For example, the same persona being consistently presented across platforms is consistent with how authentic users behave on social media. However, threat actors have also displayed this behaviour as a way to increase the perceived legitimacy of their fabricated personas (aka "backstopping").
TA16: Establish Legitimacy **T0144: Persona Legitimacy Evidence** **T0144.001: Present Persona across Platforms** Summary: This sub-technique covers situations where analysts have identified the same persona being presented across multiple platforms. Having multiple accounts presenting the same persona is not an indicator of inauthentic behaviour; many people create accounts and present as themselves on multiple platforms. However, threat actors are known to present the same persona across multiple platforms, benefiting from an increase in perceived legitimacy.
TA16: Establish Legitimacy **T0144: Persona Legitimacy Evidence** **T0144.002: Persona Template** Summary: Threat actors have been observed following a template when filling their accounts' online profiles. This may be done to enable account holders to quickly present themselves as a real person with a targeted persona. For example, an actor may be instructed to create many fabricated local accounts for use in an operation using a

Funded by the European Union

template of “[flag emojis], [location], [personal quote], [political party] supporter” in their account’s description.

Associated Techniques and Sub-techniques
T0143.002: Fabricated Persona: The use of a templated account biography in a collection of accounts may be an indicator that the personas have been fabricated.

T0145: Establish Account Imagery

T0145: Establish Account Imagery allows analysts to document account imagery often adopted by actors.

Link to technique introduction

Proposed DISARM Additions
TA15: Establish Assets **T0145: Establish Account Imagery** Summary: Introduce visual elements to an account where a platform allows this functionality (e.g. a profile picture, a cover photo, etc). Threat Actors who don’t want to use pictures of themselves in their social media accounts may use alternate imagery to make their account appear more legitimate.
TA15: Establish Assets **T0145: Establish Account Imagery** **T0145.001: Copy Account Imagery** Summary: Account imagery copied from an existing account. Analysts may use reverse image search tools to try to identify previous uses of account imagery (e.g. a profile picture) by other accounts. Threat Actors have been known to copy existing accounts’ imagery to impersonate said accounts, or to provide imagery for unrelated accounts which aren’t intended to impersonate the original assets’ owner. Associated Techniques and Sub-techniques T0143.003: Impersonated Persona: Actors may copy existing accounts’ imagery in an attempt to impersonate them. T0143.004: Parody Persona: Actors may copy existing accounts’ imagery as part of a parody of that account.
TA15: Establish Assets **T0145: Establish Account Imagery** **T0145.002: AI-Generated Account Imagery** Summary: AI Generated images used in account imagery. An influence operation might flesh out its account by uploading account imagery (e.g. a profile picture), increasing its perceived legitimacy. By using an AI-generated picture for this purpose, they are able to present themselves as a real person without compromising their own identity, or risking detection by taking a real person’s existing profile picture. Associated Techniques and Sub-techniques T0086.002: Develop AI-Generated Images (Deepfakes): Analysts should use this sub-technique to document use of AI generated imagery used to support narratives.

Funded by the European Union

TA15: Establish Assets
T0145: Establish Account Imagery
T0145.003: Animal Account Imagery
Summary: Animal used in account imagery.

An influence operation might flesh out its account by uploading a profile picture, increasing its perceived authenticity.

People sometimes legitimately use images of animals as their profile pictures (e.g. of their pets), and threat actors can mimic this behaviour to avoid the risk of detection associated with stealing or AI-generating profile pictures (see T0145.001: Copy Account Imagery and T0145.002: AI-Generated Account Imagery).

This Technique is often used by Coordinated Inauthentic Behaviour accounts (CIBs). A collection of accounts displaying the same behaviour using similar account imagery can indicate the presence of CIB.

TA15: Establish Assets
T0145: Establish Account Imagery
T0145.004: Scenery Account Imagery
Summary: Scenery or nature used in account imagery.

An influence operation might flesh out its account by uploading account imagery (e.g. a profile picture), increasing its perceived authenticity.

People sometimes legitimately use images of scenery as their profile picture, and threat actors can mimic this behaviour to avoid the risk of detection associated with stealing or AI-generating profile pictures (see T0145.001: Copy Account Imagery and T0145.002: AI-Generated Account Imagery).

This Technique is often used by Coordinated Inauthentic Behaviour accounts (CIBs). A collection of accounts displaying the same behaviour using similar account imagery can indicate the presence of CIB.

TA15: Establish Assets
T0145: Establish Account Imagery
T0145.005: Illustrated Character Account Imagery
Summary: A cartoon/illustrated/anime character used in account imagery.

An influence operation might flesh out its account by uploading account imagery (e.g. a profile picture), increasing its perceived authenticity.

People sometimes legitimately use images of illustrated characters as their profile picture, and threat actors can mimic this behaviour to avoid the risk of detection associated with stealing or AI-generating profile pictures (see T0145.001: Copy Account Imagery and T0145.002: AI-Generated Account Imagery).

This Technique is often used by Coordinated Inauthentic Behaviour accounts (CIBs). A collection of accounts displaying the same behaviour using similar account imagery can indicate the presence of CIB.

TA15: Establish Assets
T0145: Establish Account Imagery
T0145.006: Attractive Person Account Imagery
Summary: Attractive person used in account imagery.

An influence operation might flesh out its account by uploading account imagery (e.g. a profile picture), increasing its perceived authenticity.

Funded by the European Union

Pictures of physically attractive people can benefit threat actors by increasing attention given to their posts.

People sometimes legitimately use images of attractive people as their profile picture, and threat actors can mimic this behaviour to avoid the risk of detection associated with stealing or AI-generating profile pictures (see T0145.001: Copy Account Imagery and T0145.002: AI-Generated Account Imagery).

This Technique is often used by Coordinated Inauthentic Behaviour accounts (CIBs). A collection of accounts displaying the same behaviour using similar account imagery can indicate the presence of CIB.

Associated Techniques and Sub-techniques
T0097.109: Romantic Suitor Persona: Accounts presenting as a romantic suitor may use an attractive person in their account imagery.
T0104.002: Dating App: Analysts can use this sub-technique for tagging cases where an account has been identified as using a dating platform.

TA15: Establish Assets
T0145: Establish Account Imagery
T0145.007: Stock Image Account Imagery
Summary: Stock images used in account imagery.

Stock image websites produce photos of people in various situations. Threat Actors can purchase or appropriate these images for use in their account imagery, increasing perceived legitimacy while avoiding the risk of detection associated with stealing or AI-generating profile pictures (see T0145.001: Copy Account Imagery and T0145.002: AI-Generated Account Imagery).

Stock images tend to include physically attractive people, and this can benefit threat actors by increasing attention given to their posts.

This Technique is often used by Coordinated Inauthentic Behaviour accounts (CIBs). A collection of accounts displaying the same behaviour using similar account imagery can indicate the presence of CIB.

T0085.008: Machine Translated Text

During work on the update there was the opportunity to introduce the requested T0085.008: Machine Translated Text to T0085: Develop Text-Based Content.

Proposed DISARM Additions
TA06: Develop Content **T0085: Develop Text-Based Content** **T0085.008: Machine Translated Text** Summary: Text which has been translated into another language using machine translation tools, such as AI.

Funded by the European Union

Removed

T0099.003: Impersonate Existing Organisation
T0099.004: Impersonate Existing Media Outlet
T0099.005: Impersonate Existing Official
T0099.006: Impersonate Existing Influencer

These sub-techniques' parent technique *T0099: Impersonate Existing Entity* was reworked into the new sub-technique *T0143.004: Impersonated Persona* of *T0143: Persona Legitimacy*, a technique which houses different assertions one can make about the legitimacy of a given persona (e.g. *T0143.001: Authentic Persona, T0143.002: Fabricated Persona, T0143.003: Parody Persona,* and *T0143.004: Impersonated Persona*)

Each of *T0099: Impersonate Existing Entity*'s sub-techniques need to be removed as a consequence. The new approach of separating out the legitimacy of a persona from the type of persona is better for the future of the framework, and is more practical for analysts. This is particularly true in cases when the impersonation isn't initially obvious; we can easily imagine quickly spotting an impersonation of the BBC (or other well known news outlets), but impersonations of smaller outlets may be harder to identify right away. By separating the assertion of "impersonation" and "type of persona", we enable analysts to nail down the overt "what persona is this" element before investigating "is that persona legitimate".

Removed Technique	Comments
TA16: Establish Legitimacy **T0099: Impersonate Existing Entity** **T0099.003: Impersonate Existing Organisation** Summary: A situation where a threat actor styles their online assets or content to mimic an existing organisation. This can be done to take advantage of peoples' trust in the organisation to increase narrative believability, to smear the organisation, or to make the organisation less trustworthy.	*T0099.003: Impersonate Existing Organisation can now be asserted using T0143.004: Impersonated Persona and T0097.200: Institutional Persona*
TA16: Establish Legitimacy **T0099: Impersonate Existing Entity** **T0099.004: Impersonate Existing Media Outlet** Summary: A situation where a threat actor styles their online assets or content to mimic an existing media outlet. This can be done to take advantage of peoples' trust in the outlet to increase narrative believability, to smear the outlet, or to make the outlet less trustworthy.	*T0099.004: Impersonate Existing Media Outlet can now be asserted using T0143.004: Impersonated Persona and T0097.202: News Outlet Persona*
TA16: Establish Legitimacy	*T0099.005: Impersonate Existing Official can now be*

Funded by the European Union

T0099: Impersonate Existing Entity **T0099.005: Impersonate Existing Official** Summary: A situation where a threat actor styles their online assets or content to impersonate an official (including government officials, organisation officials, etc).	*asserted using T0143.004: Impersonated Persona and T0097.111: Government Official Persona*
TA16: Establish Legitimacy **T0099: Impersonate Existing Entity** **T0099.006: Impersonate Existing Influencer** Summary: A situation where a threat actor styles their online assets or content to impersonate an influencer or celebrity, typically to exploit users' existing faith in the impersonated target.	*T0099.006: Impersonate Existing Influencer can now be asserted using T0143.004: Impersonated Persona and T0097.100: Individual Persona*

T0009: Create Fake Experts

This update introduced new ways to track the use of expert personas (T0097.108: Expert Persona), and enables analysts to document the presence of experts who are 'fake' (T0143.002: Fabricated Persona), authentic, parody, or an impersonation (T0143.001: Authentic Persona, T0143.003: Parody Persona, and T0143.004: Impersonated Persona). These new techniques make T0009: Create Fake Experts redundant, so it has been removed from the framework.

Removed Technique	Comments
TA16: Establish Legitimacy **T0009: Create Fake Experts** Summary: Stories planted or promoted in computational propaganda operations often make use of experts fabricated from whole cloth, sometimes specifically for the story itself.	*T0009: Create Fake Experts can now be asserted using T0143.002: Fabricated Persona and T0097.108: Expert Persona*

T0009.001: Utilise Academic/Pseudoscientific Justifications

For DISARM to be a useful resource we need to have good cross-coder reliability; if two different analysts have different interpretations of a technique, then the data they produce becomes unreliable. This technique requires analysts to assess whether a narrative has included "academic" or "pseudoscientific" justifications as part of its messaging. This is difficult for analysts to confidently assert, and different analysts may have different interpretations of what is or isn't pseudoscience - particularly as the technique did not elaborate on its criteria in its summary.

There may be value in having techniques which document different logical fallacies, or which allow analysts to assert that narratives are misrepresented as being scientifically sound. However, DISARM is not producing a narrative framework at this time. With the removal of T0009: Create Fake Experts, the decision was made to deprecate this technique, rather than introducing new narrative focused techniques for documenting the use of "pseudoscience".

Funded by the European Union

Removed Technique	Comments
TA16: Establish Legitimacy **T0009: Create Fake Experts** **T0009.001: Utilise Academic/Pseudoscientific Justifications** Summary: Utilise Academic/Pseudoscientific Justifications	-

T0142: Fabricate Grassroots Movement

In DISARM Red V1.4, the existing technique *T0099: Prepare Assets Impersonating Legitimate Entities* was reworked into *T0099: Impersonate Existing Entity.* While making this change, its sub-technique *T0099.001: Astroturfing* was upgraded to T0142: Fabricate Grassroots movements. In this update T0142 is being removed, given its overlap with the newly introduced *T0097.103: Activist Persona, T0097.104: Hacktivist Persona, and T0097.208: Social Cause Persona* in *T0097: Present Persona*, and with Y in *T0143: Persona Legitimacy.*

It may be useful in the future to allow analysts to make the assertion that many fabricated activist, hacktivist, or social cause personas have been deployed to give the impression of a fabricated grassroots movement. DISARM is currently prioritising introducing more granular techniques to help analysts document lower level behaviour, but may look to add more “inferred” techniques to the framework in the future.

Removed Technique	Comments
TA16: Establish Legitimacy **T0142: Fabricate Grassroots Movement** Summary: This technique, sometimes known as "astroturfing", occurs when an influence operation disguises itself as a grassroots movement or organisation that supports operation narratives. Astroturfing aims to increase the appearance of popular support for an evolving grassroots movement in contrast to "Utilise Butterfly Attacks", which aims to discredit an existing grassroots movement. This Technique was previously called Astroturfing, and used the ID T0099.001	*T0142: Fabricate Grassroots Movement can now be asserted using T0143.002: Fabricated Persona and T0097.103: Activist Persona, T0097.104: Hacktivist Persona, or T0097.208: Social Cause Persona.*

Reworked

T0097: Create Personas

Create Personas was reworked into Presented Personas. It now helps analysts focus on documenting what type of persona is presented.

Funded by the European Union

Previous Technique	Updated Technique
TA16: Establish Legitimacy **T0097: Create Personas** Summary: Creating fake people, often with accounts across multiple platforms. These personas can be as simple as a name, can contain slightly more background like location, profile pictures, backstory, or can be effectively backstopped with indicators like fake identity documents.	**TA16: Establish Legitimacy** **T0097: Present Persona** Summary: This Technique contains different types of personas commonly taken on by threat actors during influence operations. Analysts should use T0097's Subtechniques to document the type of persona which an account is presenting. For example, an account which describes itself as being a journalist can be tagged with T0097.102: Journalist Persona. Personas presented by individuals include: - T0097.101: Local Persona - T0097.102: Journalist Persona - T0097.103: Activist Persona - T0097.104: Hacktivist Persona - T0097.105: Military Personnel Persona - T0097.106: Recruiter Persona - T0097.107: Researcher Persona - T0097.108: Expert Persona - T0097.109: Romantic Suitor Persona This Technique also houses institutional personas commonly taken on by threat actors: - T0097.201: Local Institution Persona - T0097.202: News Outlet Persona - T0097.203: Fact Checking Organisation Persona - T0097.204: Think Tank Persona - T0097.205: Business Persona - T0097.206: Government Institution Persona - T0097.207: NGO Persona - T0097.208: Social Cause Persona By using a persona, a threat actor is adding the perceived legitimacy of the persona to their narratives and activities.

T0097.001: Produce Evidence for Persona

Create Persona's existing sub-technique T0097.001: Present Evidence for Persona was reworked into a top level technique which houses behaviours which analysts might observe which evidences the persona's legitimacy.

Funded by the European Union

Previous Technique	Updated Technique
TA16: Establish Legitimacy **T0097: Create Personas** **T0097.001: Produce Evidence for Persona** Summary: People may produce evidence which supports the persona they are deploying (T0097) (aka "backstopping" the persona). This Technique covers situations where evidence is developed or produced as part of an influence operation to increase the perceived legitimacy of a persona used during IO, including creating accounts for the same persona on multiple platforms. The use of personas (T0097), and providing evidence to improve people's perception of one's persona (T0097.001), are not necessarily malicious or inauthentic. However, sometimes people use personas to increase the perceived legitimacy of narratives for malicious purposes. This Technique was previously called Backstop Personas.	**TA16: Establish Legitimacy** **T0144: Persona Legitimacy Evidence** Summary: This Technique contains behaviours which might indicate whether a persona is legitimate, a fabrication, or a parody. For example, the same persona being consistently presented across platforms is consistent with how authentic users behave on social media. However, threat actors have also displayed this behaviour as a way to increase the perceived legitimacy of their fabricated personas (aka "backstopping").

T0099: Impersonate Existing Entity

Impersonation is now asserted using the sub-technique T0143.004: Impersonated Persona.

Previous Technique	Updated Technique
TA16: Establish Legitimacy **T0099: Impersonate Existing Entity** Summary: An influence operation may prepare assets impersonating existing entities (both organisations and people) to further conceal its network identity and add a layer of legitimacy to its operation content. Existing entities may include authentic news outlets, public figures, organisations, or state entities. Users will more likely believe and less likely fact-check news from recognisable sources rather than unknown sites. An influence operation may use a wide variety of cyber techniques to impersonate a legitimate entity's website or social media account. This Technique was previously called Prepare Assets Impersonating Legitimate Entities.	**TA16: Establish Legitimacy** **T0143: Persona Legitimacy** **T0143.004: Impersonated Persona** Summary: Threat actors may impersonate existing individuals or institutions to conceal their network identity, add legitimacy to content, or harm the impersonated target's reputation. This Technique covers situations where an actor presents themselves as another existing individual or institution. This Technique was previously called Prepare Assets Impersonating Legitimate Entities and used the ID T0099. **Associated Techniques and Subtechniques** **T0097: Present Persona:** Analysts can use the Subtechniques of T0097: Present Persona to categorise the type of impersonation.

Funded by the European Union

	For example, a document developed by a threat actor which falsely presented as a letter from a government department could be documented using T0085.004: Develop Document, T0143.004: Impersonated Persona, and T0097.206: Government Institution Persona. **T0145.001: Copy Account Imagery:** Actors may take existing accounts' profile pictures as part of their impersonation efforts.

T0099.002: Spoof/Parody Account/Site

Parody is now asserted using the sub-technique T0143.003: Parody Persona.

Previous Technique	Updated Technique
TA16: Establish Legitimacy **T0099: Impersonate Existing Entity** **T0099.002: Spoof/Parody Account/Site** Summary: An influence operation may prepare assets impersonating legitimate entities to further conceal its network identity and add a layer of legitimacy to its operation content. Users will more likely believe and less likely fact-check news from recognisable sources rather than unknown sites. Legitimate entities may include authentic news outlets, public figures, organisations, or state entities.	**TA16: Establish Legitimacy** **T0143: Persona Legitimacy** **T0143.003: Parody Persona** Summary: Summary: Parody is a form of artistic expression that imitates the style or characteristics of a particular work, genre, or individual in a humorous or satirical way, often to comment on or critique the original work or subject matter. People may present as parodies to create humour or make a point by exaggerating or altering elements of the original, while still maintaining recognizable elements. The use of parody is not an indication of inauthentic or malicious behaviour; parody allows people to present ideas or criticisms in a comedic or exaggerated manner, softening the impact of sensitive or contentious topics. Because parody is often protected as a form of free speech or artistic expression, it provides a legal and social framework for discussing controversial issues. However, parody personas may be perceived as authentic personas, leading to people mistakenly believing that a parody account's statements represent the real opinions of a parodied target. Threat actors may also use the guise of parody to spread campaign content. Parody personas may disclaim that they are operating as a parody, however this is not always the case, and is not always given prominence. This Technique was previously called Spoof/Parody Account/Site, and used the ID T0099.002. **Associated Techniques and Subtechniques**

Funded by the European Union

	T0097: Present Persona: Analysts can use the sub-techniques of T0097: Present Persona to categorise the type of parody. For example, an account presenting as a parody of a business could be documented using T0097.205: Business Persona and T0143.003: Parody Persona. **T0145.001: Copy Account Imagery:** Actors may take existing accounts' profile pictures as part of their parody efforts.

Small Changes

Associated Techniques and Sub-techniques have been added for some existing Red framework items.

T0085.001: Develop AI-Generated Text

Previous Technique	Updated Technique
TA06: Develop Content **T0085: Develop Text-Based Content** **T0085.001: Develop AI-Generated Text** Summary: AI-generated texts refers to synthetic text composed by computers using text-generating AI technology. Autonomous generation refers to content created by a bot without human input, also known as bot-created content generation. Autonomous generation represents the next step in automation after language generation and may lead to automated journalism. An influence operation may use read fakes or autonomous generation to quickly develop and distribute content to the target audience.	**TA06: Develop Content** **T0085: Develop Text-Based Content** **T0085.001: Develop AI-Generated Text** Summary: AI-generated texts refers to synthetic text composed by computers using text-generating AI technology. Autonomous generation refers to content created by a bot without human input, also known as bot-created content generation. Autonomous generation represents the next step in automation after language generation and may lead to automated journalism. An influence operation may use read fakes or autonomous generation to quickly develop and distribute content to the target audience. **Associated Techniques and Subtechniques:** **T0085.008: Machine Translated Text:** Use this subtechnique when AI has been used to generate a translation of a piece of text

T0086.002: Develop AI-Generated Images

Previous Technique	Updated Technique
TA06: Develop Content **T0086: Develop Image-Based Content** **T0086.002: Develop AI-Generated Images (Deepfakes)** Summary: Deepfakes refer to AI-generated falsified	**TA06: Develop Content** **T0086: Develop Image-Based Content** **T0086.002: Develop AI-Generated Images (Deepfakes)** Summary: Deepfakes refer to AI-generated falsified

Funded by the European Union

photos, videos, or soundbites. An influence operation may use deepfakes to depict an inauthentic situation by synthetically recreating an individual's face, body, voice, and physical gestures.	photos, videos, or soundbites. An influence operation may use deepfakes to depict an inauthentic situation by synthetically recreating an individual's face, body, voice, and physical gestures. **Associated Techniques and Subtechniques:** **T0145.002: AI-Generated Account Imagery:** Analysts should use this sub-technique to document use of AI generated imagery in accounts' profile pictures or other account imagery.

T0104.002: Dating App

Previous Technique	Updated Technique
TA07: Select Channels and Affordances **T0104: Social Networks** **T0104.002: Dating App** Summary: "Dating App" refers to any platform (or platform feature) in which the ostensive purpose is for users to develop a physical/romantic relationship with other users. Threat Actors can exploit users' quest for love to trick them into doing things like revealing sensitive information or giving them money. Examples include Tinder, Bumble, Grindr, Facebook Dating, Tantan, Badoo, Plenty of Fish, hinge, LOVOO, OkCupid, happn, and Mamba.	**TA07: Select Channels and Affordances** **T0104: Social Networks** **T0104.002: Dating App** Summary: "Dating App" refers to any platform (or platform feature) in which the ostensive purpose is for users to develop a physical/romantic relationship with other users. Threat Actors can exploit users' quest for love to trick them into doing things like revealing sensitive information or giving them money. Examples include Tinder, Bumble, Grindr, Facebook Dating, Tantan, Badoo, Plenty of Fish, hinge, LOVOO, OkCupid, happn, and Mamba. **Associated Techniques and Sub-techniques** **T0097.109: Romantic Suitor Persona:** Analysts can use this sub-technique for tagging cases where an account presents itself as seeking a romantic or physical connection with another person.

Funded by the European Union

Annex 3 - DISARM Version 1.6 Release Notes

DISARM Red V1.6 Patch Notes

In this update we're continuing our efforts to improve interoperability between Meta's Online Operations Kill Chain (MOOKC) expanding the roster of non-content assets (e.g. social media pages, accounts, websites, etc) which can be logged using the DISARM Red Framework, while improving the framework's structure.

Design Philosophy

A More Approachable Framework

People have told us that DISARM Red is overwhelming and confusing to use. One of the ways we're trying to address this issue is by adding new clear top-level techniques which group similar sub-techniques. Tidying away framework items behind clear categorising techniques helps people find what they're looking for, and gives us a good foundation for expanding our roster of sub-techniques in the future.

New techniques introduced include:

- TA07: Select Channels and Affordances
    - T0151: Digital Community Hosting Asset
    - T0152: Digital Content Hosting Asset
    - T0153: Digital Content Delivery Asset
    - T0154: Digital Content Creation Asset
- TA15: Establish Assets
    - T0146: Account Asset
    - T0147: Software Asset
    - T0148: Financial Instrument
    - T0149: Online Infrastructure

These 8 techniques allow us to introduce 74 new sub-techniques which analysts can use to track assets used in influence operations, while keeping the top-level framework approachable. We are also merging 9 existing techniques as sub-techniques of these new techniques.

We originally designed this update to include a new tactic *("TA19: Digital Assets")* which would house each of the new techniques and sub-techniques, however we decided that this was too much of a change to the framework where the new techniques could reasonably fit into the existing tactics.

Saying One Thing at a Time

Some analysts using DISARM have given feedback that they feel uncomfortable using the existing techniques *T0090: Create Inauthentic Accounts* and *T0007: Create Inauthentic Social Media Pages and Groups*, because they include the assertion that the assets are inauthentic.

To address this, we're continuing our design goal of having techniques assert one thing at a time; the new T0146: Account Asset, *T0151.002: Online Community Group*, and *T0151.003: Online Community Page* remove assertions of inauthenticity, and are exclusively used to document the assets which are present in a potential incident.

We're introducing two new techniques which help analysts assert information about an asset which they can use alongside new framework items to further describe them:

- TA07: Select Channels and Affordances
    - T0155: Gated Asset
- TA15: Establish Assets
    - T0150: Asset Origin

For example, a compromised website can be documented using (T0150.005: Compromised Asset, *T0151.002: Online Community Group*); accounts created in bulk can be documented using *(T0150.008: Bulk Created Asset, T0146: Account Asset);* Facebook groups which require manual approval before users join can be documented using *(T0155.003: Approval Gated Asset, T0151.002: Online Community Group)*.

Introducing New Content

Throughout this introduction to new techniques and sub-techniques, examples from new incidents introduced alongside the update have been drawn out to provide real-world examples of their usage, some of which make reference to extremist behaviour online.

For ease of reference, a version of the DISARM Navigator has been produced which exclusively contains items introduced in the 1.6 update. To use the Navigator, right click a technique or sub-technique and select "view technique" to open its full description, and list of incidents.

TA07: Select Channels and Affordances

TA07 gains five new Techniques; T0151: Digital Community Hosting Asset, T0152: Digital Content Hosting Asset, T0153: Digital Content Delivery Asset, T0154: Digital Content Creation Asset, T0155: Gated Asset.

Digital Community Hosting, Content Hosting, Content Delivery, and Content Creation Assets

TA07 sees four new techniques introduced which categorise various types of online assets based on their purported primary use case; content creation, hosting, or delivery, and community hosting.

These techniques are intended to guide users who are searching for sub-techniques to document specific types of online assets, and to provide a generic technique for analysts to use when they come across an online asset which doesn't fit one of the sub-techniques.

For example, *T0151.001: Social Media Platform* is housed in T0151: Digital Community Hosting Asset, even though social media platforms also have features that host, deliver and help create content. They contain both online platforms (such as *T0151.010: Community Forum Platform* (e.g. Reddit)) and assets users can make on those

platforms (such as *T0151.011: Community Sub-Forum* (e.g. Subreddits)). The following table lists all the sub-techniques.

Sub #	T0151: Digital Community Hosting Asset	T0152: Digital Content Hosting Asset	T0153: Digital Content Delivery Asset	T0154: Digital Content Creation Asset
1	Social Media Platform	Blogging Platform	Email Platform	AI LLM Platform
2	Online Community Group	Blog Asset	Link Shortening Platform	AI Media Platform
3	Online Community Page	Website Hosting Platform	Shortened Link Asset	
4	Chat Platform	Website Asset	QR Code Asset	
5	Chat Community Server	Paste Platform	Online Advertising Platform	
6	Chat Room	Video Platform	Content Recommendation Algorithm	
7	Chat Broadcast Group	Audio Platform	Direct Messaging	
8	Microblogging Platform	Live Streaming Platform		
9	Legacy Online Forum Platform	Software Delivery Platform		
10	Community Forum Platform	File Hosting Platform		
11	Community Sub-Forum	Wiki Platform		
12	Image Board Platform	Subscription Service Platform		
13	Question and Answer Platform			
14	Comments Section			
15	Online Game Platform			
16	Online Game Session			
17	Dating Platform			

The sub-techniques provide examples of popular platforms or features for each technique. They provide descriptions of the platforms or features which the sub-technique covers, and list associated techniques and sub-techniques. Let's look at T0151.008: Microblogging Platform as an example of this structure:

Technique	Summary
T0151.008: Microblogging Platform	Examples of popular Microblogging Platforms include Threads, Bluesky, Mastodon, QQ, Tumblr, TikTok, and X (formerly Twitter). **Microblogging** platforms allow users to create **Accounts**, which they can configure to present themselves to other platform users. This typically involves **Establishing Account Imagery,** and **Presenting a Persona**. **Accounts** on **Microblogging Platforms** are able to post short-form text content alongside media. Content posted to the platforms is aggregated into different feeds and presented to the user. Typical feeds include content posted by other **Accounts** which the user follows, and content promoted by the platform's proprietary **Content Recommendation Algorithm.** Users can also search or use hashtags to discover new content. Mastodon is an open-source decentralised software which allows anyone to create their own **Microblogging Platforms** that can communicate with other platforms within the "fediverse" (similar to how different email platforms can send emails to each other). Meta's **Threads** is a **Microblogging Platform** which can interact with the fediverse.

Also provided are a list of reports which have been tagged as containing the sub-technique, with relevant quotes pulled out, and examples of how the sub-technique (and other DISARM techniques and sub-techniques) map to the report. Continuing the example from *T0151.008: Microblogging Platform:*

Report	Mapping
Teen who hacked Joe Biden and Bill Gates' Twitter accounts sentenced to three years in prison	*An 18-year-old hacker who pulled off a huge breach in 2020, infiltrating several high profile Twitter accounts to solicit bitcoin transactions, has agreed to serve three years in prison for his actions.* *Graham Ivan Clark, of Florida, was 17 years old at the time of the hack in July, during which he took over a number of major accounts including those of Joe Biden, Bill Gates and Kim Kardashian West.* *Once he accessed them, Clark tweeted a link to a bitcoin address and wrote "all bitcoin sent to our address below will be sent back to you doubled!" According to court documents, Clark made more than $100,000 from the scheme, which his lawyers say he has since returned.* *Clark was able to access the accounts after convincing an employee at Twitter he worked in the company's information technology department, according to the Tampa Bay Times.* In this example a threat actor gained access to Twitter's customer service portal through

	social engineering (T0146.004: Administrator Account Asset, T0150.005: Compromised Asset, T0151.008: Microblogging Platform), which they used to take over accounts of public figures (T0146.003: Verified Account Asset, T0143.003: Impersonated Persona, T0150.005: Compromised Asset, T0151.008: Microblogging Platform). The threat actor used these compromised accounts to trick their followers into sending bitcoin to their wallet (T0148.009: Cryptocurrency Wallet).

This section of the document will provide introductions to newly introduced techniques and surface examples of their usage in reporting.

T0155: Gated Asset

Also introduced to TA07 is T0155: Gated Asset, a technique which houses different sub-techniques analysts can use to assert that an asset is not openly accessible, and what form of gating is in place:

- *T0155.001: Password Gated Asset*
- *T0155.002: Invite Gated Asset*
- *T0155.003: Approval Gated Asset*
- *T0155.004: Geoblocked Asset*
- *T0155.005: Paid Access Asset*
- *T0155.006: Subscription Access Asset*

Report	Mapping
Gaming the System: The Use of Gaming-Adjacent Communication, Game and Mod Platforms by Extremist Actors	In this report, researchers look at online platforms commonly used by people who play videogames, looking at how these platforms can contribute to radicalisation of gamers: *Gamer Uprising Forums (GUF) [is an online discussion platform using the classic forum structure] aimed directly at gamers. It is run by US Neo-Nazi Andrew Anglin and explicitly targets politically right-wing gamers. This forum mainly includes antisemitic, sexist, and racist topics, but also posts on related issues such as esotericism, conspiracy narratives, pro-Russian propaganda, alternative medicine, Christian religion, content related to the incel- and manosphere, lists of criminal offences committed by non-white people, links to right-wing news sites, homophobia and trans-hostility, troll guides, anti-leftism, ableism and much more. Most noticeable were the high number of antisemitic references. For example, there is a thread with hundreds of machine-generated images, most of which feature openly antisemitic content and popular antisemitic references. Many users chose explicitly antisemitic avatars. Some of the usernames also provide clues to the users' ideologies and profiles feature swastikas as a type of progress bar and indicator of the user's activity in the forum.* *The GUF's front page contains an overview of the forum, user statistics, and so-called "announcements". In addition to advice-like references, these feature various expressions of hateful ideologies. At the time of the exploration, the following could be read there: "Jews are the problem!", "Women should be raped", "The Jews are going to be required to return stolen property", "Immigrants will have to be physically removed", "Console gaming is for n******" and "Anger is a womanly emotion". New users have to prove*

	themselves in an area for newcomers referred to in imageboard slang as the "Newfag Barn". Only when the newcomers' posts have received a substantial number of likes from established users, are they allowed to post in other parts of the forum. It can be assumed that this will also lead to competitions to outdo each other in posting extreme content. However, it is always possible to view all posts and content on the site. In any case, it can be assumed that the platform hardly addresses milieus that are not already radicalised or at risk of radicalisation and is therefore deemed relevant for radicalisation research. However, the number of registered users is low (typical for radicalised milieus) and, hence, the platform may only be of interest when studying a small group of highly radicalised individuals. Gamer Uprising Forum is a legacy online forum, with access gated behind approval of existing platform users (T0155.003: Approval Gated Asset, T0151.009: Legacy Online Forum Platform)

TA15: Establish Assets

T0146: Account Asset

T0146: Account Asset is a new technique which analysts can use to assert that an actor has an account on a particular platform. For example, an account on the eCommerce platform Etsy can be documented using *(T0146: Account Asset, T0148.007: eCommerce Platform)*.

Report	Mapping
Facebook Is Being Flooded With Gross AI-Generated Images of Hurricane Helene Devastation	*As families desperately seek to find missing loved ones and communities grapple with immeasurable losses of both life and property in the wake of [2024's] Hurricane Helene, AI slop scammers appear to be capitalizing on the moment for personal gain.* *A Facebook account called "Coastal Views" usually shares calmer AI imagery of nature-filled beachside scenes. The account's banner image showcases a signpost reading "OBX Live," OBX being shorthand for North Carolina's Outer Banks islands.* *But starting this weekend, the account shifted its approach dramatically, as first flagged by a social media user on X.* *Instead of posting "photos" of leaping dolphins and sandy beaches, the account suddenly started publishing images of flooded mountain neighborhoods, submerged houses, and dogs sitting on top of roofs.* *But instead of spreading vital information to those affected by the natural disaster, or at the very least sharing real photos of the destruction, the account is seemingly trying to use AI to cash in on all the attention the hurricane has been getting.* *The account links to an Etsy page for a business called" OuterBanks2023," where somebody who goes by "Alexandr" sells AI-generated prints of horses touching snouts with sea turtles, Santa running down the shoreline with a reindeer, and sunsets over ocean waves.*

	A Facebook page which presented itself as being associated with North Carolina which posted AI generated images changed to posting AI generated images of hurricane damage after Hurricane Helene hit North Carolina (T0151.003: Online Community Page, T0151.001: Social Media Platform, T0115: Post Content, T0086.002: Develop AI-Generated Images (Deepfakes), T0068: Respond to Breaking News Event or Active Crisis). The account included links (T0122: Direct Users to Alternative Platforms) to an account on Etsy, which sold prints of AI generated images (T0146: Account Asset, T0148.007: eCommerce Platform).

Its sub-techniques contain different types of accounts analysts may want to document:

- *T0146.001: Free Account Asset*
- *T0146.002: Paid Account Asset*
- *T0146.003: Verified Account Asset*
- *T0146.004: Administrator Account Asset*
- *T0146.005: Lookalike Account ID*
- *T0146.006: Open Access Platform*
- *T0146.007: Automated Account Asset*

Sub-technique *T0146.006: Open Access Platform* is used to document platforms which don't require accounts for users to access their content or features.

T0147: Software Asset

T0147: Software Asset is a technique which houses different types of software which actors can acquire for use in an influence operation:

- T0147.001: Game Asset
- T0147.002: Game Mod Asset
- T0147.003: Malware Asset
- T0147.004: Mobile App Asset

T0148: Financial Instrument

T0148: Financial Instrument houses platforms and capabilities related to financial transactions:

- T0148.001: Online Banking Platform
- T0148.002: Bank Account Asset
- T0148.003: Payment Processing Platform
- T0148.004: Payment Processing Capability
- T0148.005: Subscription Processing Capability
- T0148.006: Crowdfunding Platform
- T0148.007: eCommerce Platform
- T0148.008: Cryptocurrency Exchange Platform
- T0148.009: Cryptocurrency Wallet

T0148.004: Payment Processing Capability and *T0148.005: Subscription Processing Capability* allow analysts to document the ability for actors to process payments without specifying that they have an account on a payment processing platform (such as PayPal).

Report	Mapping
Coordinated Facebook Pages Designed to Fund a White Supremacist Agenda	This report examines the white nationalist group Suavelos' use of Facebook to draw visitors to its website without overtly revealing their racist ideology. This section of the report looks at the Suavelos website, and the content it links out to. *In going back to Suavelos' main page, we also found: A link to a page on a web shop: alabastro.eu; A link to a page to donate money to the founders through Tipee and to the website through PayPal; [and] a link to a private forum that gathers 3.000 members: oppidum.suavelos.eu;* Suavelos linked out to an online store which it controlled (T0152.004: Website Asset, T0148.004: Payment Processing Capability), and to accounts on payment processing platforms PayPal and Tipee (T0146: Account Asset, T0148.003: Payment Processing Platform). The Suavelos website also hosted a private forum (T0151.009: Legacy Online Forum Platform, T0155: Gated Asset), and linked out to a variety of assets it controlled on other online platforms: accounts on Twitter (T0146: Account Asset, T0151.008: Microblogging Platform), YouTube (T0146: Account Asset, T0152.006: Video Platform), Instagram and VKontakte (T0146: Account Asset, T0151.001: Social Media Platform).

T0149: Online Infrastructure

T0149: Online Infrastructure collates technical assets which support actors in their influence operation:

- T0149.001: Domain Asset
- T0149.002: Email Domain Asset
- T0149.003: Lookalike Domain
- T0149.004: Redirecting Domain Asset
- T0149.005: Server Asset
- T0149.006: IP Address Asset
- T0149.007: VPN Asset
- T0149.008: Proxy IP Address Asset
- T0149.009: Internet Connected Physical Asset

Report	Mapping
Charming Kitten Updates POWERSTAR with an	*The target of the recently observed [highly targeted spearphishing attack by "Charming Kitten", a hacker group attributed to Iran] had published an article related to Iran. The publicity appears to have garnered the attention of Charming Kitten, who subsequently created an email address to impersonate a reporter of an Israeli media organization in*

InterPlanetary Twist	*order to send the target an email. Prior to sending malware to the target, the attacker simply asked if the target would be open to reviewing a document they had written related to US foreign policy. The target agreed to do so, since this was not an unusual request; they are frequently asked by journalists to review opinion pieces relating to their field of work.* *In an effort to further gain the target's confidence, Charming Kitten continued the interaction with another benign email containing a list of questions, to which the target then responded with answers. After multiple days of benign and seemingly legitimate interaction, Charming Kitten finally sent a "draft report"; this was the first time anything opaquely malicious occurred. The "draft report" was, in fact, a password-protected RAR file containing a malicious LNK file. The password for the RAR file was provided in a subsequent email.* In this example, threat actors created an email address on a domain which impersonated an existing Israeli news organisation, and a reporter who worked there (T0097.102: Journalist Persona, T0097.202: News Outlet Persona, T0143.003: Impersonated Persona, T0149.003: Lookalike Domain, T0149.002: Email Domain Asset) in order to convince the target to download a document containing malware (T0085.004: Develop Document, T0147.003: Malware Asset).

Documentation

Mapping to MOOKC

This update enables 28 new MOOKC mappings, and lays groundwork for 39 more centring on malicious use of assets. Analysts looking to map to MOOKC can refer to documentation provided here.

Mapping to DISARM

As mentioned above, this update fixed issues raised with *T0090: Create Inauthentic Accounts* with the new T0146: Account Asset. With each improvement to the DISARM Framework we document all changes made to existing framework items, and collate them in this document for ease of reference, including how to map content tagged with previous DISARM techniques to the updated framework.

Annex 4 - DISARM Red Retrospective and Futrospective

DISARM Red Retrospective and Futrospective

In this post we reflect on the updates DISARM has released as part of the Horizon funded ADAC.IO project over the last year; what went right, what still needs work, and improvements to what comes next.

We start by revisiting users' feedback on the existing DISARM Framework (here called 'DISARM v1'). We then assess how we addressed this feedback in the approach taken for updates published as part of the ADAC.IO project (here called 'DISARM v1.5'), both its successes and challenges. We finish by introducing our plans for DISARM v2.

DISARM v1 Approach

Early in the ADAC.IO project we gathered feedback from DISARM users and stakeholders. We published a full report on the results (which you can read here), but we've summarised key points below:

Why People Used DISARM

Sensemaking

Analysts like referring to the Framework to think about what kind of behaviours they could be looking for in a potential incident, and developing methods to counter these commonly exhibited behaviours.

Data Sharing

Users saw the potential for gathering data on incidents at scale using DISARM, using this to better understand threat actors, and to facilitate data sharing across defenders. Tagging reports with DISARM TTPs provide a sense of what is happening across different reports, and data sharing can enable users to faster identify when incidents are connected as part of a campaign as well as answer research questions about the prevalence of different tactics.

Users' Pain Points

Prior to and during the project, we gathered feedback from users on what they wanted to improve about DISARM. Overall, users fell into two categories; time-pressured users who wanted an easy tagging experience, and advanced users who wanted a larger, more objective framework. They told us:

The Framework is Too Complicated

A common piece of feedback was that the framework was too complicated. There were several issues that contributed to this:

Confusing Structure: The Framework was structured so that it progressed left-to-right through stages an actor may take during an influence operation (referred to as the "Kill Chain"), meaning it opened with things like planning objectives, strategy, and analysing a target audience. While thinking about this is useful for sensemaking, analysts almost never see operational planning, making this a poor introduction to the Framework.

Poor User Experience: The website which hosted the Framework prioritised function over form. This worked for our highly technical users, but needlessly adds complexity for an already complex topic.

Unclear Tactics and Techniques: Some Tactics and Techniques were not easy to understand just by their name, and didn't explain themselves in the description. Some Techniques' names inferred a different meaning to their descriptions, which added further confusion. Inter-coder reliability is essential for successful data sharing, so Techniques shouldn't leave any room for misinterpretation.

Too Many Techniques: Some users told us they were overwhelmed by the size of the Framework, and wanted a much more streamlined collection of key behaviours.

There are Significant Gaps in the Framework

Running contrary to users who claimed the Framework needed slimming down, other analysts raised concerns that gaps in the Framework meant they couldn't document critical aspects of campaigns they'd uncovered.

Judgements in Techniques

Some analysts felt uncomfortable using Techniques which contained judgmental language, such as "Create Inauthentic Accounts", because they felt unable to assert that an account was "inauthentic", and the criteria to determine an account's 'inauthenticity' wasn't clearly defined in the Technique's description. Furthermore, the inclusion of "Inauthentic" prevented documenting 'authentic' accounts, belonging to real people, amplifying false or misleading information.

DISARM Updates Too Slowly

Threat actors are constantly releasing new methods of manipulation. DISARM needs to update quickly, not only to plug existing gaps, but to handle the new behaviours. For DISARM to succeed as a common language for sharing data about FIMI, it needs to be able to quickly adapt to threat actors' innovations.

DISARM Doesn't Collaborate with the Community Enough

Many DISARM users have suggestions for how to refine existing Techniques—and some have drafted full Techniques they would like to see added to the Framework. DISARM needs to work better with the community to address this feedback.

DISARM v1.5 Approach

One of the core goals of the ADAC.IO project was to increase interoperability between the DISARM Framework and other counter-FIMI tools. This process began by reviewing the Meta's Online Operations Kill Chain ("MOOKC"—their version of a TTP Framework), identifying behaviours which couldn't be described using DISARM, and adding them to the DISARM Framework. This increased interoperability, and plugged gaps in the Framework.

Change in Design Philosophy

Issues with Techniques which Assert Many Things at Once

In both MOOKC and DISARM techniques describe behaviours in aggregate, saying multiple things at once (we call these "Aggregate Techniques"). For example, DISARM's "Create Inauthentic Accounts" identifies (1) someone has created (2) an account (3) which is being used 'inauthentically'. Meta took the same approach—for example, MOOKC's "1.5.1 Impersonating news website" identifies (1) an impersonation (2) of a news outlet (3) which is hosted on a website.

These Aggregate Techniques contributed to some of the issues raised by users:

Missing Techniques: Aggregate Techniques tie several unique observations together to create a specific description of a behaviour. This creates frustrating cases where *most* of a Technique applies, but not all of it ("they're definitely creating an account, but I can't tell if it's 'inauthentic'").

Unclear Tactics and Techniques: There are also inconsistencies in what Aggregate Techniques aggregate. MOOKC sometimes describes the asset used in an impersonation (1.5.1 Impersonating news website), sometimes they don't (1.4.1 Impersonating researcher or think tanker). This lack of consistency makes it hard to feel confident about how elements of an influence operation should be documented.

Too Many Techniques: Expanding coverage using Aggregate Techniques would exponentially increase the size of the Framework. Focusing just on expanding coverage of personas (how an asset is presenting itself—a journalist, a news website, a researcher, etc) and the persona's legitimacy (is it fake, an impersonation, etc), if we wanted to be able to document a new type of persona (e.g. assets presenting themselves as Wellness Influencers) then a unique technique would be introduced for each category of legitimacy (e.g. Fake Wellness Influencer, Impersonated Wellness Influencer, etc). Following Meta's example and including information about the asset used would further propagate the number of Techniques in the framework (e.g. Fake Wellness Influencer Account, Fake Wellness Influencer Website, Fake Wellness Influencer Facebook Page, Impersonated Wellness Influencer Account, etc). Given the existing feedback that the Framework had too many Techniques, this approach was not feasible.

Techniques Designed to be Combined

New additions to the DISARM framework would enable individual observations to be made about a potential FIMI incident; instead of "Fake Wellness Influencer Account", we would have "Fake", "Wellness Influencer", and "Account". Analysts can use a combination of Techniques to describe the unique situation they are faced with.

Reviewing the Success of v1.5 Updates

DISARM produced four updates under the v1.5 design philosophy, each increasing in scope and impact. In working on a larger fifth update, we identified issues which necessitated another change in our approach. In this section we look at what went well, and what challenges we faced, in the v1.5 updates:

What Went Well

Efficiently Closed Gaps in the Framework

Providing many unique Techniques for analysts to combine was the most efficient way to expand DISARM's coverage without propagating the number of Techniques in the framework. For example, the Persona-focused v1.5.5 update introduced just 38 new Techniques which, when combined, provided the same level of coverage as 176 Aggregate Techniques.

Optional Judgements

Analysts who felt unable to use Techniques such as "Create Inauthentic Accounts" because of the inclusion of "Inauthentic" could now simply combine "Create" and "Accounts", without assessing a persona's authenticity. Analysts who felt able to make an assessment about the persona's legitimacy could do so by adding in e.g. "Impersonated", "Fake", "Parody", or "Authentic".

Clearer Techniques

With Techniques only describing one thing at a time, defining them became a lot simpler. To further increase Technique clarity, we introduced over 300 real-world examples of new Techniques being used in the form of Incidents, which displayed under new Techniques in the framework, pulling out quotes from reports and explaining how the quote demonstrated that behaviour.

What Challenges did we face?

Throughout the year the DISARM team became more familiar with the processes required to make changes, and as such each update became incrementally larger as we were able to fit in more improvements; where v1.5.5 introduced 38 new Techniques, v1.5.6 introduced 82, and the unreleased v1.5.7 update was slated to add over 100 new Techniques. As these updates became larger, the issues with the new approach became more apparent.

In this section we lay out the challenges we faced, before moving on to how we plan to address them in the following section.

The Framework got Even More Complicated

Confusing Technique Names: In early ADAC.IO updates we prioritised giving Techniques short names for efficiency. In some cases this meant analysts would have to read the Technique's description to fully understand what it documented.

As we progressed through updates, we realised analysts preferred longer, more descriptive Technique names which got the message across without requiring a click through to read more.

Too Many Techniques: Even though adding Techniques which say one thing at a time was an approach which filled gaps in the Framework while minimising the total Techniques added, it still meant a lot of additions to the Framework, which was the opposite of what some people were asking for.

Changing Existing Techniques: In early ADAC.IO updates we made changes to existing Techniques to more clearly get across what they were meant to document, including updating their name (while maintaining the same 'identifier'). People didn't like this; Techniques which had left room for interpretation changed to document a more specific behaviour, which didn't always match analysts' interpretations of the Technique (and therefore what they had used it to document).

In later updates we instead fully removed and replaced any Techniques which we felt required large changes, meaning reports tagged with previously existing Techniques were not impacted.

People Liked Aggregate Techniques and We Removed Them: When we introduced new Techniques which said one thing at a time, we removed old Aggregate Techniques which were covered by the new Techniques. For example, "Create Inauthentic Accounts" was removed and replaced with "Newly Created Asset", "Fabricated Persona", and "Account Asset". This was intended to ensure that Techniques in the framework were mutually exclusive; we didn't want people documenting the same behaviour with different Techniques.

However, people who wanted a smaller Framework of essential behaviours saw the kinds of Techniques they had memorised being removed, and replaced with highly granular Techniques. These users are least likely to want to learn a new system, or to trawl through documentation, so they were particularly impacted by these changes. This is also true of people who used the Framework for sensemaking, who liked browsing Aggregate Techniques to help them think about the kinds of behaviours they might see in a campaign.

We have plans for how DISARM v2 can support Aggregate Techniques in a more targeted way, which is better for sensemaking, and for people who want a streamlined collection of behaviours (detailed later in this post).

Combining Techniques wasn't a Well Established Workflow: Before the ADAC.IO updates, DISARM had some Techniques which needed to be paired with other Techniques to deliver meaningful information. For example, "Post Content" by itself doesn't convey much, but when paired with techniques like "'Harass", "Incentivise Sharing", or "Bait Influencer", we get a more useful description of an action. However, most items in the framework were Aggregate Techniques, which delivered useful information without needing to associate them with other Techniques; Analysts were not commonly combining Techniques in their work.

To help address this issue, DISARM tagged third party reporting and published them alongside updates, both as a way to show Techniques' use in the real world, but also to demonstrate how Techniques could be applied together in reporting. However, these tagged reports were only displayed at the bottom of Technique description pages, which were hard to access (as discussed below).

When working on the 1.5.7 update, it became clear that we couldn't rely exclusively on DISARM tagged reports as the only support guiding analysts on how to pair DISARM Techniques.

Difficulties Accessing the Updates

A New, More Complicated User Experience: The v1 DISARM Framework was hosted on a Herokuapp instance which the DISARM team is not able to update for technical reasons. All v1.5 updates were published to a

Navigator hosted on Github. The Navigator provides more tools for users to browse the Framework, but is harder for non-technical users to use, and requires a few too many click throughs to launch.

Informing People about Updates: Without the ability to update the Herokuapp instance, we could not easily inform our users of new updates released. It was announced in a blog post and linked to from the website, but this didn't get through to enough people. DISARM doesn't have a complete list of users, so we didn't have a good way to inform people about the new platform (you can help us fix that here).

While we wrote detailed Patch Notes describing each update, and provided documentation for mapping newly introduced Techniques to removed ones, these similarly suffered from lack of visibility.

Unaddressed Issues

Still a Confusing Structure: With the new requirement for associating Techniques, the confusing structure of the Framework became more impactful as Techniques in different "phases" of an operation needed to be paired.

DISARM Still Updates too Slowly: We released an update approximately once every three months, which wasn't quick enough. Updates steadily increased in size and impact, but this growth led to the v1.5.7 update being considered too much of a change to publish. Instead of releasing the update, the decision was made to use 1.5.7 work as the basis for a new approach in DISARM v2.

DISARM Still Doesn't Work Well with the Community: While mapping to Meta's Online Operations Kill Chain was a good approach for quickly identifying gaps in the framework and achieving ADAC.IO's goal of increasing interoperability, in focusing on the MOOKC, we centered the work of a non-DISARM user. We still included some changes based on user feedback, but some were to be addressed in the unreleased v1.5.7 update, leaving users disappointed. Further, the community rightly expected more interaction than occasionally addressing their feedback in a patch; they want more collaboration, more communication, and more output.

v1.5 Summary

Changing our approach to having Techniques observe one thing at a time helped address users' issues with applying judgemental Techniques, and helped us expand DISARM's coverage while minimising the Framework's expansion.

However, this approach was not without issue. The Aggregate Techniques we removed were familiar to many users, and their absence made the Framework harder to navigate, especially for those who preferred a smaller, more conceptual set of behaviours rather than granular, composable parts. We needed a way to still support users who benefited from Aggregate Techniques.

DISARM v2 Approach

In this section we detail our plans for DISARM v2, and how they address challenges we faced in DISARM v1.5 updates. Throughout we highlight **calls to action,** where we solicit your feedback on our plans. We want DISARM to be the best tool for documenting and sharing data on influence operations and online harms. You can use this

form to let us know what you think of our proposed changes, and tell us how we can best support you in your usage of DISARM.

The v2 approach can be broken down into three key areas; improved **updates**, improved **support**, and improved **framework**.

Improved Updates

v2 updates will be focused on particular topic areas in what we are calling Themed Collections. Collections can focus on harms (e.g. Technology Facilitated Gender-Based Violence), actors (e.g. Doppelganger, Portal Kombat, Online Platforms), DISARM user types (e.g. Fact Checkers), countermeasures (e.g. DSA) or topics (e.g. Elections). Development of Themed Collections involves improving the framework's ability to document its area of focus (introducing new Techniques, and refining existing Techniques where required), and producing supporting materials (more detail in the next section).

How this addresses v1.5 Issues

Producing Themed Collections targeted on specific topics helps address:

- **DISARM Still Updates too Slowly:** While there are still behind-the-scenes technical issues for us to address to reduce time *between* updates, development of Themed Collections focusing on different topics can be worked on concurrently, helping us deliver *more* improvements in the same timeframe.
- **DISARM Still Doesn't Work Well with the Community:** By focusing updates on particular themes, it makes it easier for us to reach out to relevant organisations to ask for their insights, expertise, and feedback. This approach also helps us better work with members of the community who reach out with requests for tactical Framework improvements through a "DISARM User Community" Collection.
- **Informing People about Updates:** Working with members of the community to develop updates pertaining to their area of focus creates awareness of DISARM updates most relevant to them. Supporting materials developed as part of a Themed Collection also provide better change documentation than the previous Patch-Notes-per-update solution.

  Call to Action: What topic would you like a DISARM update to focus on? What smaller, tactical improvements would you like to see in a DISARM User Community Themed Collection?

Improved Support

Themed Collections include a variety of supporting materials which we are calling DISARM Playbooks; resources designed to make it easier for people to use DISARM. Examples of the different types of Playbooks currently in development include:

- **Key Techniques:** A list of Techniques relevant to the Collection's focus, supporting time pressured analysts find the items they need in a Framework of many Techniques.
- **Quick Reference Guide:** Common behaviours observed in that Collection's focus, mapped to DISARM. This guide helps us fill the gap left by "Aggregate Techniques", providing an easy way to document larger behaviours (e.g. amplifying a narrative on an inauthentic news site) using paired DISARM Techniques, and providing examples of how Techniques can be paired together to document threat behaviours.

- **Detailed Tagging Support:** Guides on how to make changes to the templated Aggregate Techniques provided in the **Quick Reference Guide** to adapt to the unique situations analysts may face. This helps analysts whose intelligence requirements necessitate documenting an observed incident in greater detail.
- **Tagged Reports:** A compilation of third party reports around the Collection's focus, tagged by members of the DISARM team. These real-world examples of Techniques' usage help ensure analysts are on the same page about what a Technique documents, and gives more references for how DISARM Techniques can be paired.
- **Decision Trees:** A step-by-step guide on how to make tagging decisions when facing a potential incident within the Collection's focus. This reduces complexity by helping describe to analysts the different choices they can make in pairing Techniques to describe an incident.

These resources currently exist as documentation which analysts can refer to. However, they have been designed to support the production of features in a future digital platform which hosts the DISARM Framework. For example, the collection of Key Techniques could be used to provide prebuilt Framework filters for analysts to quickly focus down on behaviours relevant to their line of investigation, once we have the resources to begin building out the necessary tech stack.

How this addresses v1.5 Issues

Producing supporting materials alongside updates in the form of DISARM Playbooks helps address:

- **Combining Techniques wasn't a Well Established Workflow:** Playbooks provide many examples of how DISARM Techniques can be combined to deliver a more detailed picture of potential incidents.
- **People Liked Aggregate Techniques and We Removed Them:** The **Quick Reference Guide** provides the same "Aggregate Technique" experience analysts liked in the v1 Framework by providing simple ways to document larger behaviours.
- **Still a Confusing Structure:** Playbooks provide another way to access and benefit from DISARM Techniques in a way which is less overwhelming than delving straight into the full framework.
- **DISARM Still Doesn't Work Well with the Community:** Unique Playbooks can be developed to support the particular niche the Themed Collection covers. DISARM can work with members of the community not only to improve the Framework's ability to document their area of focus, but also to identify unique pain points which can be addressed through support materials bespoke to that topic.

> **Call to Action:** We want to support our users in applying the DISARM Framework. What kind of materials would you like DISARM to produce to achieve this goal? What resources would you like to see us develop?

Improved Framework

A Change in Backwards Compatibility

Updates which remove Techniques raise several issues for analysts; losing those they are familiar with and use regularly is a frustrating experience, and can invalidate reports tagged using these deprecated Techniques.

In v1.5 updates we removed some existing Techniques from the framework, while ensuring we maintained the ability to describe removed behaviours using newly introduced Techniques, and producing documentation for how

analysts could do this (including updating tagged reports with the new Techniques). However, hard-to-find documentation was not a good enough solution, especially for organisations who potentially faced retagging a huge amount of work tagged on previous versions of the Framework. It's also the case that some of the removed Techniques definitions were unclear, and that analysts' interpretations of what they documented differed from the replacement Techniques we offered for them.

In v2, we will *deprecate* Techniques instead of *removing* them. This means, alongside our existing documentation mapping deprecated to new Techniques, we will make previous versions of the Framework accessible (including those before the v2 update). This will provide analysts with access to historic Techniques, without hosting them alongside their replacements in the core Framework, meaning that incidents documented using old Techniques would provide the same information as it did when originally tagged by analysts, without need for rework them every time DISARM updates.

In the short term this will be achieved by making public the data files used to produce each version of the Framework. Our long term goal is to provide a more user-friendly way to access old versions of the Framework within the same UI as the live version of the Framework—with the caveat that we are still assessing technical feasibility for this feature, and prioritising it against other features users are asking for.

How this addresses v1.5 Issues

- **People Liked Aggregate Techniques and We Removed Them:** Ideally people will feel comfortable applying newly introduced Techniques—however they will have the option to still use deprecated Techniques if this better suits their workflow.

  Call to Action: DISARM will continue to change as we expand and refine coverage of known threat behaviours, and as actors develop new methods for propagating online harms. How can we best support you in adapting to updates to the Framework?

A New Organising Principle

We're changing the way the Framework is organised to have descriptive categories designed to help our users find the Techniques they're looking for, rather than organising Techniques into unique Tactical goals under a Kill Chain structure.

Some of our users really like that Techniques are organised under the Kill Chain (i.e. stages of an influence operation, such as planning objectives, target research, narratives creation, establishing assets, publishing and amplifying content), but others find it really confusing (the first thing they see when they begin documenting an incident is assessments of the actor's strategic goal, their objectives, and audience analysis, actions they rarely have visibility of), and some argue that while a rigid Kill Chain structure works in cybersecurity, influence operations are much more fluid; behaviours can be exhibited in multiple Kill Chain stages, and threat actors can move back and forth along the Kill Chain as their operation evolves.

Furthermore, in a Kill Chain structure behaviours are tied to a specific stage of the Kill Chain, which limits analysts in how they can describe incidents. For example, because "T0117: Attract Traditional Media" sits in "TA09: Deliver Content", when an analyst observes an asset attempting to attract traditional media, they are also asserting that the tactical goal of this action is to deliver content, rather than to maximise exposure (TA17: Maximise Exposure) or to

establish legitimacy (TA16: Establish Legitimacy). By untying Techniques from stages of the Kill Chain, we can give analysts the option to decide for themselves which stage of the Kill Chain a behaviour occurs in.

How this addresses v1.5 Issues

- **Still a Confusing Structure:** This new structure has been designed to help users find what they're looking for in a large Framework, while also progressing the objective of separating out assessments of an actor's objective from their observed behaviours.

Splitting Assessments from Observations

We're splitting out Assessments (like "What is the actor's strategic objective", "Who is sponsoring this actor?", "What stage of the breakout scale is this IO on?") and Observations (like "An undisclosed deepfake was posted", "An account was impersonating a journalist", "A network worked together to publish slight variations of the same text") into different Frameworks.

This separation helps us provide a more logical experience for analysts; when they want to document what they are seeing, they use the DISARM Red Observations Framework. When they want to build on their Observations, they use the DISARM Red Assessments Framework.

> **Call to Action:** How do you feel about splitting Observations (things analysts observe about a potential incident, like what was posted, or who was posting it) from Assessments (things analysts assess about a potential incident, like who was behind it, or what their objective was)?

Reworking Terminology

Inspired by the MITRE ATT&CK Framework, the DISARM Framework used a lot of terminology from the cybersecurity field. These terms were useful to help people 'get' DISARM quickly, but as the Framework has evolved, these concepts became less and less applicable to it, and may instead help people quickly get the wrong impression.

For example, "TTP" stands for "Tactics, Techniques, and Procedures". People refer to "DISARM TTPs" as a shorthand for "things in the DISARM Framework" even though we don't support any Ps. Earlier we discussed moving away from using the Kill Chain as an organising structure for the DISARM Red Framework, which enables analysts to make their own assessment about what the tactical goal of an observed behaviour is. This change would mean that the columns which organise the Framework are no longer "Tactics".

This would leave us with DISARM Ts—but calling items in the DISARM Red Framework "Techniques" limits what can be documented. For example, it would be useful for analysts to be able to use DISARM to document their assessment of which stage of The Breakout Scale an IO is on, but "Stage of the Breakout Scale" is not a "Technique" used by actors. Further, calling framework items Techniques again coaches all analysts' observations in an implicit assessment that the behaviour is being exhibited as a way to achieve a specific tactical goal.

You can see our current proposal for new naming conventions to replace "DISARM Red Framework", "Techniques", and "TTPs" below:

Existing Terminology → DISARM Red Framework: A Framework housing Techniques used by threat actors in influence operations

New Terminology → DISARM Red Framework: A Framework used to describe potential IO incidents, made up of the DISARM Red Observations Framework and the DISARM Red Assessments Framework.
New Terminology → DISARM Red Observations Framework: A collection of Observations analysts can make about a potential IO incident
New Terminology → DISARM Red Assessments Framework: A collection of Assessments analysts can make about a potential IO

Existing Terminology → Techniques: Techniques used by threat actors to achieve tactical goals in an influence operations

New Terminology → Observations: Things analysts can observe when documenting potential IO incidents
New Terminology → Assessments: Assessments analysts make about an actor behind a potential IO incident
New Terminology → Aggregates: Combinations of Observations and Assessments which describe behaviours exhibited by Threat Actors in influence operations

Existing Terminology → TTPs: Tactics, Techniques, and Procedures used by threat actors in influence operations

New Terminology → Framework Items: Individual items on within the DISARM Red Observations or Assessments Frameworks

> **Call to Action:** We're aware people are fond of existing terminology, and that it can be frustrating to have to get used to name changes, but we hope we've made the case for why it's important to make this change. Do you agree with the new proposed terms? Would you have any suggestions for changes?

A Preview of the Observations Framework

The DISARM Red Observations Framework contains three key top level categories; **Assets, Actions, and Content,** each with their own sub categories, which contain Observations analysts can make about a potential incident. The current iteration of the beta v2 Observations Framework has the following categories:

- Assets
  - —Accounts, Software, and Infrastructure
  - —Digital Platforms and Communities
  - —Asset Details
  - —Asset Persona
- Actions
  - —Publishing and Amplifying Actions
  - —Asset Management
  - —Interact with Platform and Community
  - —Offline Actions
  - —Actions Taken by Platforms
- Content
  - —Content Acquisition

—Content Details
—Content Narrative
—Content Action

We believe that these descriptive categories will better guide analysts in finding the Observations they're looking for; an issue repeatedly raised with DISARM's v1 organising principle. You can view a beta version of the DISARM Red Observations Framework hosted in the DISARM Navigator here, or you can view an experimental platform hosting a prototype of the v2 DISARM Framework here.

How this addresses v1.5 Issues

- **Too Many Techniques:** Splitting Observations from Assessments reduces the size of each Framework. Introducing a more approachable organising principle for the Framework also helps make Observations and Assessments easier to find, which should reduce the impact of the number Framework Items on its usability.
- **Still a Confusing Structure:** Splitting Observations and Assessments helps sort Framework items based on their intended purpose, reducing analyst overload.

What's Next

We've outlined our plan for DISARM v2, and why we think it improves on DISARM v1 and v1.5. In this section, we talk about moving forward with the plan.

DISARM v2 Development

Work on getting DISARM v2 ready for release is ongoing. We're preparing the core v2 Observations and Assessments Framework, along with a collection of Themed Collections to publish alongside them, including a Technology Facilitated Gendered Violence Collection, a Fact Checking Collection, and a Portal Kombat Collection.

In the meantime, you can use this form to let us know you'd like to be notified about the release of DISARM v2, and to give us feedback on anything you've read here (or on improvements you'd like to see to the Framework). We want to make sure DISARM is the best tool it can be for documenting influence operations and online harms, and we want to hear from you on how we can make that happen.

DISARM v2 Rollout

There are still issues unaddressed by the DISARM v2 plans outlined above.

A New Platform

DISARM currently uses open-source technology developed by MITRE ATT&CK to house the DISARM Red Framework. This was a cost-effective solution which let us offset development work to developers in an organisation with the skills and resources to handle it (and who were kind enough to let us use the fruits of their

labour). However, this tool was designed for highly technical cybersecurity analysts, which doesn't match the DISARM user profile. Plus, the DISARM Framework is already larger than the MITRE ATT&CK Framework, and is likely to expand further as threat actors develop new methods, or as emerging technologies enable new threat behaviours. DISARM will need more tools to help users find what they're looking for than those available in the MITRE ATT&CK Navigator.

The need for a bespoke platform to access the DISARM Framework is clear. We've begun work on scoping requirements for such a platform, balancing our wildest dreams with what can be attained with limited funding. This platform will be designed with the goal of simplifying the use of DISARM, while also removing much of the existing friction we face in updating the DISARM Framework from a technical perspective (resulting in faster updates).

> **Call to Action:** What would you like included in a platform developed by DISARM? Are there features you've seen in other platforms you'd like included? What problems do you need help solving?

Conclusion

Thank you for taking the time to hear about our plans for improving DISARM. We hope that these changes will make it easier for the defender community to collaborate, share data, and fight back against people who try to use manipulation to undermine democracy.

Annex 5 - DISARM Playbooks Overview

DISARM Playbooks

This post explores the development of DISARM Playbooks and how they support analysts working within the DISARM framework.

What are DISARM Playbooks?

DISARM Playbooks are resources designed to help people use DISARM within given topic areas. This is an intentionally broad definition as we explore how we can support the range of ways people use DISARM, and what is achievable with available resources.

Most DISARM Playbooks are currently different types of documentation for analysts to refer to as they use DISARM, but have been produced with the goal of using Playbooks' content to inform development of technical interventions (once DISARM has the resources to develop said interventions). These pieces of documentation come together to make a Playbook on a topic. Playbooks can focus on a variety of topics, including narrative / focus area (e.g. Technology Facilitated Gender-Based Violence (TFGBV)), defender type (e.g. Fact Checker), or threat actor (e.g. Portal Kombat).

The following are types of resources which have been developed in the last year as part of Playbooks focusing on those topics:

Key Techniques

There are a lot of different behaviours exhibited in influence operations, and as such DISARM has a lot of items in the Framework, which people have told us is overwhelming. The Key Techniques resource addresses this issue by providing a list of Techniques which are—key—to documenting behaviours in a given topic area.

Below are descriptions of how this resource has been produced over the course of different updates to the DISARM Framework, with links to each Playbook iteration:

- **TFGBV: Key Techniques— DISARM v2 Alpha:** This first iteration of Key Techniques pulled out full descriptions of Techniques which document harms associated with TFGBV.
- **Fact Checking: Key Techniques—DISARM v1.7:** We produced short descriptions of Techniques which were key for Fact Checkers and collated them in this document. Provision of short descriptions made the resource more concise, and better achieved the goal of providing an overview of Key Techniques.

Tagged Reports

DISARM applies TTPs to public third party reports, giving analysts real-world examples of how they have been used, and examples of how reports can be augmented using DISARM. This gives analysts the ability to double check their understanding of TTPs against DISARM's work, improving inter-coder reliability, and DISARM's ease of use.

Below are descriptions of how this resource has been produced over the course of different updates to the DISARM Framework, with links to each Playbook iteration:

- **TFGBV: Tagged Reports—DISARM v2 Alpha:** The first iteration of Tagged Reports identified at least one third party report for each Key Technique, and provided a DISARM-augmented version of each. It also provided a Procedure (then called 'Aggregate') from each report.
- **Fact Checking: Tagged Reports—DISARM v1.7:** There were more Key Techniques for Fact Checkers than for TFGBV, so a resource collating a tagged report for each would not be very usable. This resource instead highlighted reports for the most commonly occurring behaviours (the KEY Key Techniques), and for Techniques which required more nuance / analysis to apply. Analysts can still refer to the >100 DISARM-augmented third party reports introduced as part of the update within the framework—this resource just serves to highlight particularly useful / demonstrative ones.

Prefabricated Procedures

In DISARM v2, Procedures are DISARM Observations which are aggregated using DISARM's standardised Procedure formula (you can read more about this here). The Prefabricated Procedures resource provides examples of commonly seen Procedures in a given topic area, or key behaviours which can be slotted into Procedures as an analyst sees fit. The goal of this resource is to help time-pressured analysts quickly benefit from DISARM's Procedures system without having to learn it in-depth.

An example of this resource is provided below:

- **TFGBV: Prefabricated Procedures—DISARM v2 Alpha**: This first iteration of Prefabricated Procedures was developed before 'Observations as Procedures' was a fully fleshed out concept, and attempted to minimise the use of the Procedure system with the goal of reducing potential overwhelm of readers. Future versions of this resource will lean more into the idea of DISARM Procedures, both how and why they should be used.

Tagging Support

We have explored ways to support analysts making real-time tagging decisions. This has included discussion of 'decision trees'—but something like this would be better suited to a technical intervention, which DISARM does not currently have the resources to produce. In the meantime, Playbook resources have been produced which support analysts in making choices about what to tag in an incident, examples of which are provided below:

- **TFGBV: Tagging Support—DISARM v2 Alpha**: This first draft of a Playbook in DISARM combined Key Techniques, Tagged Reports, and Prefabricated Procedures into one large document. Initial feedback was that this was too much in one document—and as such it was broken down into separate resources (detailed above).
- **Portal Kombat: Tagging Support—DISARM v2 Alpha:** This resource ties together an introduction to Portal Kombat, some Prefabricated Procedures based on common Portal Kombat behaviours, and provides an example tagging scenario to support the use of Procedures. It also provides support for mapping coordinated behaviour between Portal Kombat and other well known influence operations, showing how Playbooks can provide different types of support depending on their area of focus.

- **Fact Checking: Tagging Support—DISARM v1.7:** This resource focused on helping Fact Checkers operationalise DISARM with two core sections. The first looked at DISARM as a way to answer intelligence questions, and used this to help analysts TTPs they would apply to their work (thinking about what information was important to them, and what they had the capability and capacity to uncover). The second section provided a guide on how to identify behaviours within content addressed by Fact Checkers, organising TTPs based on how commonly they occurred.

Other Resources

We welcome feedback, as well as insight into systems you have implemented and would be willing to share with the community, or pain points you'd like help overcoming. Please get in touch with any feedback on DISARM Playbooks, and watch this space for further updates.

Annex 6 - DISARM TFGBV Playbook - Key Techniques

TFBGV: Key DISARM Observations

Key DISARM Observations are part of DISARM Playbooks - materials designed to support analysts on applying DISARM in a given topic area. *Key Observations* list Observations which are relevant for documenting the Playbook's topic area - if one of these appear in an incident, it is an indication the Playbook's topic may be in play.

List of Observations

- Content
  - T0176: Content Style
    - T0176.006: Anonymous Content
    - T0176.007: Sexually Explicit Content
  - T0180: Harmful Content
    - T0180.001: Unsolicited Sexual Imagery
    - T0180.003: Unsolicited Request for Sexual Imagery
    - T0180.004: Voyeuristic Content
    - T0180.006: Content Constitutes CSAM
    - T0180.007: Content Constitutes Sextortion
    - T0180.008: Abusive Content
    - T0180.009: Discriminatory Content
    - T0180.010: Intersectional Discriminatory Content
    - T0180.011: Content Depicts Offline Harm
  - T0167: Private Materials
    - T0167.001: Personally Identifiable Information
    - T0167.002: Private Intimate Imagery
  - T0166: AI-Generated Content
    - T0166.005: Deepfake Impersonation
    - T0166.006: AI-Nudified Imagery
  - T0168: Edited Content
    - T0168.005: Third Party Introduced to Content
  - T0179: Content Action
    - T0179.001: Content Threatens Action
    - T0179.005: Leak of Private Material
- Action
  - T0127: Offline Harm

T1027.001: Physical Violence
T0127.003: Physical Sexual Violence

Observations with Descriptions

Content

T0176: Content Style

T0176.006: Anonymous Content

Field	Content
Observation Description	This Observation can be used to document an analyst's assessment that content was intentionally made in such a way to hide the identity of the person or group who produced it. Actors producing content with the intent to abuse or threaten a target often do so while taking steps to reduce the ability of investigators to tie their real identity to the content they produced.
Associated Observations	**T0168.000: Edited Content:** Actors may make edits to media they've produced to anonymise the content. For example, audio may be edited to use a voice modulator to reduce the chance the speaker is recognised (T0174: Audio Content (T0168.000: Edited Content, T0176.006: Anonymous Content)). Video may be edited to blur the face of an individual (T0173: Video Content (T0168.000: Edited Content, T0176.006: Anonymous Content)). **T0180.001: Unsolicited Sexual Imagery:** It is often the case that unsolicited sexual imagery is anonymous (i.e. has been produced in such a way that identifying features of an individual (e.g. face, tattoos) are out of shot). Anonymous unsolicited sexual imagery should be documented using (T0176.006: Anonymous Content, T0180.001: Unsolicited Sexual Imagery)
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA31: Content Action > T0176: Content Style > T0176.006: Anonymous Content

T0176.007: Sexually Explicit Content

Field	Content
Observation Description	This Observation can be used to document an analyst's assessment that a piece of content is sexually explicit.
Associated Observations	**T0180.004: Voyeuristic Content:** Sexually explicit content which has been produced without the knowledge of one or more participants should be documented alongside this Observation; i.e. (T0176.007: Sexually Explicit Content, T0180.004: Voyeuristic Content). **T0180.001: Unsolicited Sexual Imagery:** Sexually explicit content delivered to an individual without solicitation or consent can be documented using this Observation; i.e. (T0180.001: Unsolicited Sexual Imagery). **T0167.002: Private Intimate Imagery:** Sexually explicit content produced consensually for a private audience should instead be documented using this Observation; i.e. (T0167.002: Private Intimate Imagery). **T0127.003: Physical Sexual Violence:** Content which depicts acts of sexual violence should instead be documented using this Observation; e.g. (T0173.000: Video Content (T0127.003: Physical Sexual Violence)). **T0168.005: Third Party Introduced to Content:** Sexually explicit images which have been edited to introduce individuals to the scene should be documented alongside this Observation (i.e. T0172.000: Image Content (T0176.007: Sexually Explicit Content, T0168.005: Third Party Introduced to Content)). **T0166.005: Deepfake Impersonation:** Sexually explicit deepfake impersonations should be documented using (T0166.005: Deepfake Impersonation, T0176.007: Sexually Explicit Content).
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA29: Content Details > T0176: Content Style > T0176.007: Sexually Explicit Content

T0180: Harmful Content

T0180.001: Unsolicited Sexual Imagery

Field	Content
Observation Description	Unsolicited Sexual Imagery (sometimes referred to as "cyberflashing") is defined as image or video content showing "a person's genitals, for the purpose of their own sexual gratification or to cause the victim humiliation, alarm or distress".

	This Observation can be used to document an analyst's assessment that an actor has delivered unwanted sexual imagery to a target.
Associated Observations	**T0176.006: Anonymous Content:** It is often the case that unsolicited sexual imagery is anonymous (i.e. has been produced in such a way that identifying features of an individual (e.g. face, tattoos) are out of shot). Anonymous unsolicited sexual imagery should be documented using (T0176.006: Anonymous Content, T0180.001: Unsolicited Sexual Imagery). **T0180.003: Unsolicited Request for Sexual Imagery:** Unsolicited requests for sexual imagery are often sent alongside unsolicited sexual imagery as an attempt to initiate a 'trade'. Unsolicited sexual imagery which requests sexual imagery in return should be documented using (T0180.003: Unsolicited Request for Sexual Imagery, T0180.001: Unsolicited Sexual Imagery).
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA31: Content Action > T0180: Harmful Content > T0180.001: Unsolicited Sexual Imagery

T0180.003: Unsolicited Request for Sexual Imagery

Field	Content
Observation Description	Unsolicited Request for Sexual Imagery is defined as any content which contains a request for the target to produce and deliver sexual imagery of themselves, colloquially referred to as 'nudes'. This Observation can be used to document an analyst's assessment that an actor has submitted an unwanted request for sexual imagery to a target.
Associated Observations	**T0180.001: Unsolicited Sexual Imagery:** Unsolicited requests for sexual imagery are often sent alongside unsolicited sexual imagery as an attempt to initiate a 'trade'. Unsolicited sexual imagery which requests sexual imagery in return should be documented using (T0180.003: Unsolicited Request for Sexual Imagery, T0180.001: Unsolicited Sexual Imagery)
Attribution Framework Mapping	Content Signal

Observation Location in Framework	P07: Content > TA31: Content Action > T0180: Harmful Content > T0180.003: Unsolicited Request for Sexual Imagery

T0180.004: Voyeuristic Content

Field	Content
Observation Description	Voyeuristic Content is defined as image or video content secretively taken of another individual for a sexual purpose. This Observation can be used to document an analyst's assessment that an image or video meets the criteria of Voyeuristic Content. Voyeuristic Content is typically produced using hidden cameras, or discreet photography Observations; using a zoom lens, or taking a photo when the target is not paying attention. Actors have been observed attempting to produce Voyeuristic Content targeting women by taking photos up their skirts (called "Upskirting") or down their tops ("Downblousing") without their knowledge or consent. Actors also secretively photograph women from afar for sexual gratification ("Creepshots"). Analysts will need to make an assessment about whether content was produced for sexual gratification, and whether the depicted individual was aware they were being filmed. Analysts may use context of the platform content was posted to, or how it was described by the actor, to make such an assessment.
Associated Observations	**T0176.007: Sexually Explicit Content:** Voyeuristic Content which depicts one or more individuals in a sexually explicit situation (such as secretively, non-consensually recorded consensual sex acts) should also be documented using this Observation i.e. (T0176.007: Sexually Explicit Content, T0180.004: Voyeuristic Content).
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA31: Content Action > T0180: Harmful Content > T0180.004: Voyeuristic Content

T0180.006: Content Constitutes CSAM

Field	Content

Observation Description	This Observation can be used to document an analyst's assessment that an actor has published content which depicts child sexual abuse (also known as CSAM). AI-Generated content depicting children in sexually explicit ways is considered to be CSAM. CSAM is defined by the U.S. Government as any visual depiction of sexually explicit conduct involving a person less than 18 years old.
Associated Observations	**T0166.006: AI-Nudified Imagery:** Actors have been observed using AI to produce nude versions of images depicting people less than 18 years old (T0166.006: AI-Nudified Imagery, T0180.006: Content Constitutes CSAM). **T0166.000: AI-Generated Content:** Actors have been observed using AI to generate sexually explicit material depicting people less than 18 years old (T0166.000: AI-Generated Content, T0180.006: Content Constitutes CSAM). **T0166.005: Deepfake Impersonation:** Actors have been observed using AI to generate sexually explicit deepfakes impersonating people less than 18 years old (T0166.000: AI-Generated Content, T0180.006: Content Constitutes CSAM)
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA31: Content Action > T0180: Harmful Content > T0180.006: Content Constitutes CSAM

T0180.007: Content Constitutes Sextortion

Field	Content
Observation Description	This Observation can be used to document an analyst's assessment that content attempts to Sextort a target. "Sexual extortion, or "sextortion," occurs when an individual has, or claims to have, a sexual image of another person and uses it to coerce a person into doing something they do not want to do". To make this assessment, analysts must confirm that messaging threatens the non-consensual publication of sexual content depicting the target, with a demand for action.
Associated Observations	**T0167.002: Private Intimate Imagery:** Actors have been observed using the threat of non-consensually publishing private intimate imagery depicting a target as a method of sextortion, which can be documented using (T0180.007: Content Constitutes Sextortion (T0167.002: Private Intimate Imagery)). **T0180.004: Voyeuristic Content:** Actors have been observed using the threat of publishing non-consensually produced sexually explicit content depicting a target as a

	method of sextortion, which can be documented using (T0180.007: Content Constitutes Sextortion (T0180.004: Voyeuristic Content, T0176.007: Sexually Explicit Content)). **T0166.005: Deepfake Impersonation:** Actors have been observed using the threat of publishing sexually explicit deepfake impersonations depicting a target as a method of sextortion, which can be documented using (T0180.007: Content Constitutes Sextortion (T0166.005: Deepfake Impersonation, T0176.007: Sexually Explicit Content)).
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA31: Content Action > T0180: Harmful Content > T0180.007: Content Constitutes Sextortion

T0180.008: Abusive Content

Field	Content
Observation Description	Abusive Content is defined as targeted negative messaging, unwanted by the recipient. This Observation can be used to document an analyst's assessment that an actor has published content with the intent of delivering an abusive, negative message to an individual, or group of individuals, To identify abusive content, analysts must make an assessment about the intent of the person publishing the content, and how it was experienced by its target. Did the account select the content they published with the intent to cause distress in their target? Is it reasonable to assume the actor would be aware that the content they published is upsetting to their target audience?
Associated Observations	**T0180.009: Discriminatory Content:** Abuse which targets protected characteristics should be documented alongside this Observation; i.e. (T0180.008: Abusive Content, T0180.009: Discriminatory Content). **T0179.001: Content Threatens Action:** Abuse which threatens violence should be documented alongside these Observations; i.e. (T0180.008: Abusive Content, T0179.001: Content Threatens Action (T0127.001: Physical Violence)).
Attribution Framework Mapping	Content Signal

Observation Location in Framework	P07: Content > TA31: Content Action > T0180: Harmful Content > T0180.008: Abusive Content

T0180.009: Discriminatory Content

Field	Content
Observation Description	Discriminatory Content is defined as harmful or negative messaging that perpetuates stereotypes, biases, or prejudices based on personal characteristics such as sexual orientation, gender, race, religion, disability, or other legally protected social identities or attributes. This content often reinforces harmful societal norms and can contribute to the marginalisation or stigmatisation of certain groups This Observation can be used to document an analyst's assessment that an actor has published content which discriminates against protected characteristics of an individual, or a demographic of individuals.
Associated Observations	**T0180.010: Intersectional Discriminatory Content:** Analysts should use T0180.010: Intersectional Discriminatory Content when content discriminates against multiple protected characteristics. **T0180.008: Abusive Content:** Discrimination accompanied by abuse (i.e. targeted negative messaging, unwanted by the recipient) should be documented alongside this Observation; i.e. (T0180.008: Abusive Content, T0180.009: Discriminatory Content). **T0179.001: Content Threatens Action:** Discriminatory content which threatens violence should be documented alongside these Observations; i.e. (T0180.009: Discriminatory Content, T0179.001: Content Threatens Action (T0127.001: Physical Violence)).
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA31: Content Action > T0180: Harmful Content > T0180.009: Discriminatory Content

T0180.010: Intersectional Discriminatory Content

Field	Content
Observation Description	Discriminatory Content is defined as harmful or negative messaging that perpetuates stereotypes, biases, or prejudices based on personal characteristics such as sexual orientation, gender, race, religion, disability, or other legally protected social identities or

	attributes. This content often reinforces harmful societal norms and can contribute to the marginalisation or stigmatisation of certain groups Intersectional Discrimination occurs when content discriminates against two or more protected characteristics. This Observation can be used to document an analyst's assessment that an actor has published content which discriminates against multiple protected characteristics of an individual, or a demographic of individuals.
Associated Observations	**T0180.009: Discriminatory Content:** Analysts should use T0180.009: Discriminatory Content when content discriminates against one protected characteristic only. **T0180.008: Abusive Content:** Discrimination accompanied by abuse (i.e. targeted negative messaging, unwanted by the recipient) should be documented alongside this Observation; i.e. (T0180.008: Abusive Content, T0180.010: Intersectional Discriminatory Content). **T0179.001: Content Threatens Action:** Intersectionally discriminatory content which threatens violence should be documented alongside these Observations; i.e. (T0180.010: Intersectional Discriminatory Content, T0179.001: Content Threatens Action (T0127.001: Physical Violence)).
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA31: Content Action > T0180: Harmful Content > T0180.010: Intersectional Discriminatory Content

T0180.011: Content Depicts Offline Harm

Field	Content
Observation Description	Content Depicts Offline Harm refers to content depicting offline harms, such as beheadings, sexual assaults, or physical attacks. In some cases, such content is shared to raise awareness of abuses committed by others, or to document human rights violations. However, perpetrators of offline harms have also been observed publishing footage of the harms they commit, using media as a tool of terror, control, or propaganda. This type of content can revictimise those targeted by circulating their suffering, often without consent, for the purposes of intimidation, ideology, or gratification.

Associated Observations	**T0127.001: Physical Violence:** Content depicting physical violence should be documented using (T0180.011: Content Depicts Offline Harm (T0127.001: Physical Violence)). **T0127.003: Physical Sexual Violence:** Content depicting physical sexual violence should be documented using (T0180.011: Content Depicts Offline Harm (T0127.003: Physical Sexual Violence)).
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA31: Content Action > T0180: Harmful Content > T0180.011: Content Depicts Offline Harm

T0167: Private Materials

T0167.001: Personally Identifiable Information

Field	Content
Observation Description	This Observation can be used to document an analyst's assessment that content contains personally identifiable information (or "PII"). PII is any data that can identify an individual or institution, such as their name, social accounts, physical address, contact details, medical data, information about their employer, or their friends or family. The non-consensual publication of another person's PII is commonly referred to as a "Dox", or "Doxxing". Actors have been observed uncovering and publishing targets' PII, leading to threats of physical or sexual violence against the target, and harassment against them or people they know. To make this assessment, analysts will need to identify types of sensitive information about a person or institution included in a piece of content.
Associated Observations	**T0179.005: Leak of Private Material:** Personally Identifiable Information published non-consensually ("Dox" or "Doxxing") should be documented alongside this Observation; i.e. (T0167.001: Personally Identifiable Information, T0179.005: Leak of Private Material). **T0179.001: Content Threatens Action:** Threats to leak Personally Identifiable Information should be documented alongside this Observation; e.g. (T0179.001: Content Threatens Action (T0156.000: Publish Content (T0167.001: Personally Identifiable Information, T0179.005: Leak of Private Material))).

Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA28: Content Acquisition > T0167: Private Materials > T0167.001: Personally Identifiable Information

T0167.002: Private Intimate Imagery

Field	Content
Observation Description	This Observation can be used to assert that content depicts intimate imagery consensually produced for a private audience. Private Intimate Imagery depicts an individual in a state of undress, engaging in sexual activity, or in other intimate situations where there is a reasonable expectation of privacy. This imagery is typically created or shared with the subject's consent for private use.
Associated Observations	**T0179.005: Leak of Private Material:** Private Intimate Imagery published non-consensually should be documented alongside this Observation; i.e. (T0167.002: Private Intimate Imagery, T0179.005: Leak of Private Material). **T0179.001: Content Threatens Action:** Threats to leak Private Intimate Imagery should be documented alongside this Observation; e.g. (T0179.001: Content Threatens Action (T0156.000: Publish Content (T0167.002: Private Intimate Imagery, T0179.005: Leak of Private Material))). **T0180.001: Unsolicited Sexual Imagery:** Sexually explicit content delivered to an individual without solicitation or consent can be documented using this Observation; i.e. (T0180.001: Unsolicited Sexual Imagery).
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA28: Content Acquisition > T0167: Private Materials > T0167.002: Private Intimate Imagery

T0166: AI-Generated Content

T0166.005: Deepfake Impersonation

Field	Content

Observation Description	This Observation can be used to assert that a piece of content is an AI-Generated deepfake impersonation. A deepfake refers to AI-generated content that artificially inserts the likeness of a real person into a new or altered media scene. These synthetic images, videos, or audio clips use machine learning models to manipulate or replace the original content, making it appear as though the target is participating in actions or situations they never actually did.
Associated Observations	**T0176.007: Sexually Explicit Content:** Sexually explicit deepfake impersonations should be documented using (T0166.005: Deepfake Impersonation, T0176.007: Sexually Explicit Content). **T0166.006: AI-Nudified Imagery:** Sexually explicit Deepfakes (T0166.005: Deepfake Impersonation, T0176.007: Sexually Explicit Content) are different to AI-Nudified Imagery in how they are produced. Sexually explicit Deepfakes use AI to insert an individual into sexually explicit content, or to generate an entirely fabricated scene involving the target, where Nudified images use AI to add nudity to non-sexual content.
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA28: Content Acquisition > T0166: AI-Generated Content > T0166.005: Deepfake Impersonation

T0166.006: AI-Nudified Imagery

Field	Content
Observation Description	This Observation can be used to document an analyst's assessment that AI was employed to generate AI-nudified imagery. AI-nudified imagery refers to the use of artificial intelligence to modify an existing image or video by digitally altering the subject to appear unclothed, typically through the application of machine learning algorithms that remove or replace clothing with simulated content. Platforms designed to nudify imagery require users to upload an image of a person fully clothed as the source content. The AI then applies algorithms to remove the clothing, effectively generating a synthetic image or video of the subject in a nude state.
Associated Observations	**T0166.003: Source Content for AI-Generation:** Images of individuals submitted to be 'nudified' are source materials. **T0166.005: Deepfake Impersonation:** Sexually explicit Deepfakes (T0166.005: Deepfake Impersonation, T0176.007: Sexually Explicit Content) are different to AI-Nudified Imagery in how they are produced. Sexually explicit Deepfakes use AI to insert an individual into

	sexually explicit content, or to generate an entirely fabricated scene involving the target, where Nudified images use AI to add nudity to non-sexual content.
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA28: Content Acquisition > T0166: AI-Generated Content > T0166.006: AI-Nudified Imagery

T0168: Edited Content

T0168.005: Third Party Introduced to Content

Field	Content
Observation Description	This Observation can be used to document cases where a third party individual or institution has been introduced to content which they weren't originally part of. To make this assessment, analysts will need to find the original unedited content, and identify individuals or institutions which appear in content published by the actor that do not appear in the original. Actors have been observed introducing individuals to images to associate them with target narratives.
Associated Observations	**T0176.007: Sexually Explicit Content:** Sexually explicit images which have been edited to introduce individuals to the scene should be documented alongside this Observation (i.e. T0172: Image Content (T0176.007: Sexually Explicit Content, T0168.005: Third Party Introduced to Content)).
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA28: Content Acquisition > T0168: Edited Content > T0168.005: Third Party Introduced to Content

T0179: Content Action

T0179.001: Content Threatens Action

Field	Content
Observation Description	This Observation can be used to document an analyst's assessment that a piece of content contains a threat of harmful action against a target, typically in an attempt to coerce the target into taking action desired by the actor. Actors have been observed threatening violence in an attempt to silence a target, removing their access to free speech.
Associated Observations	**T0180.007: Content Constitutes Sextortion:** Threats to non-consensually publish sexual material depicting a target should be documented using this Observation; i.e. (T0180.007: Content Constitutes Sextortion). **T0179.005: Leak of Private Material:** Threats to leak private materials can be documented using e.g. (T0197.001: Content Threatens Action (T0156.000: Publish Content (T0167.001: Personally Identifiable Information, T0179.005: Leak of Private Material))). **T0127.001: Physical Violence:** Threats of physical violence can be documented using (T0197.001: Content Threatens Action (T0127.001: Physical Violence)). **T0127.003: Physical Sexual Violence:** Threats of physical sexual violence can be documented using (T0197.001: Content Threatens Action (T0127.003: Physical Sexual Violence)).
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA31: Content Action > T0179: Content Action > T0179.001: Content Threatens Action

T0179.005: Leak of Private Material

Field	Content
Observation Description	This Observation can be used to assert that private materials have been non-consensually published to an audience not originally intended to have access to said materials. Analysts will need to assess whether content was intended to be accessible to the audience which now has access to it.
Associated Observations	**T0167.001: Personally Identifiable Information:** The non-consensual publication of personally identifiable information (colloquially referred to as 'doxxing') should be

	documented using (T0179.005: Leak of Private Material, T0167.001: Personally Identifiable Information). **T0167.002: Private Intimate Imagery:** The non-consensual publication of private intimate imagery (T0179.005: Leak of Private Material, T0167.001: Personally Identifiable Information).
Attribution Framework Mapping	Content Signal
Observation Location in Framework	P07: Content > TA31: Content Action > T0179: Content Action > T0179.005: Leak of Private Material

Action

T0127: Offline Harm

T1027.001: Physical Violence

Field	Content
Observation Description	This Observation can be used to document the act of committing offline violence. Physical violence refers to the intentional use of physical force to cause harm, injury, disability, or death to another person or group
Associated Observations	**T0180.011: Content Depicts Offline Harm:** Content depicting physical violence can be documented using (T0180.011: Content Depicts Offline Harm (T0127.001: Physical Violence)). **T0179.001: Content Threatens Action:** This Observation can be used in combination with T0127.003: Physical Sexual Violence to document threats of sexual violence (T0179.001: Content Threatens Action (T0127.001: Physical Violence)). **T0127.003: Physical Sexual Violence:** Physical sexual violence should be documented using (T0127.003: Physical Sexual Violence).
Attribution Framework Mapping	Behaviour Signal
Observation Location in Framework	P06: Actions > TA26: Offline Actions > T0127: Offline Harm >

	T0127.001: Physical Violence

T0127.003: Physical Sexual Violence

Field	Content
Observation Description	This Observation can be used to document the act of committing offline sexual violence. Sexual violence refers to any sexual act or attempt to obtain a sexual act by coercion, force, or manipulation, denying an individual the ability to freely consent. It includes a range of behaviors such as rape, sexual assault, unwanted sexual touching, sexual harassment, and exploitation.
Associated Observations	**T0180.011: Content Depicts Offline Harm:** Content depicting physical violence can be documented using (T0180.011: Content Depicts Offline Harm (T0127.003: Physical Sexual Violence)). **T0179.001: Content Threatens Action:** This Observation can be used in combination with T0127.003: Physical Sexual Violence to document threats of sexual violence (T0179.001: Content Threatens Action (T0127.003: Physical Sexual Violence)).
Attribution Framework Mapping	Behaviour Signal
Observation Location in Framework	P06: Actions > TA26: Offline Actions > T0127: Offline Harm > T0127.003: Physical Sexual Violence

Annex 7 - DISARM TFGBV Playbook - Tagged Reports

TFBGV: Tagged Reports

Tagged Reports are part of DISARM Playbooks - materials designed to support analysts on applying DISARM in a given topic area. *Tagged Reports* provide examples of how DISARM analysts have tagged third party reports published on the Playbook's topic focus.

Contents

Tagged Incidents

Chinese actress Jiang Mengjie praised for revealing upskirting blackmail

Field	Incident Properties
Name	Chinese actress Jiang Mengjie praised for revealing upskirting blackmail
Live URL	https://www.bbc.co.uk/news/world-asia-china-65326835
Date Published (YYYY/MM/DD)	2023/04/19
Authors	Kerry Allen

Publication	BBC News
Archive URL	https://web.archive.org/web/20230419150220/https://www.bbc.co.uk/news/world-asia-china-65326835
Summary	*Chinese actress Jiang Mengjie has been praised for sharing she was the victim of upskirting and was blackmailed over footage that circulated of her.*

Field	Incident Properties
Category	Sextortion, Voyeurism
Min Tags	T0180.004: Voyeuristic Content, T0180.007: Content Constitutes Sextortion
Inline Tag	*Chinese actress Jiang Mengjie has been praised for sharing she was the victim of upskirting and was blackmailed over footage that circulated of her.* *Upskirting - involving a device such as a camera phone to take explicit images underneath a victim's clothing without permission - often goes undetected.* *Jiang Mengjie told her eight million followers on social network Weibo the video had been filmed "many years ago"* ***(T0173.000: Video Content (T0180.004: Voyeuristic Content, T0170.003: Historic Content))***. *[...]* *Jiang further revealed that she had begun receiving private messages* ***(T0156.011: Send Message (T0153.007: Direct Messaging))*** *blackmailing her over the footage "saying that they would send the video to major film and TV companies and brands, and ruin the rest of my life"* ***(T0180.007: Content Constitutes Sextortion (T0179.001: Content Threatens Action (T0156.011: Send Message (T0173.000: Video Content (T0180.004: Voyeuristic Content, T0170.003: Historic Content)))))***. *[Here, a screenshot of messages is shown. In it, a user sends the messages "my patience is limited", "Sister, don't be so nervous. I'm just asking for money."* ***(T0179.008: Content Solicits Action (T0162.001: Make Payment))*** *"You have ignored me for so long, it seems you can afford the contract breach of contract, advertising breach of contract, and your acting career?]"* *[...]* *She told her followers: "as a public figure, maybe I can make more people pay attention to such vicious incidents by taking a stand.* *"It is not our fault that we have been secretly photographed. Our lives should not be affected by this kind of thing."*
Aggregate Text	A direct message which sextorts the target with the threat to release old voyeuristic footage unless payment is made
Aggregate Brackets	(T0156.011: Send Message (T0153.007: Direct Messaging), T0171.000: Text Content (T0180.007: Content Constitutes Sextortion (T0179.001: Content Threatens Action (T0156.011: Send Message (T0173.000: Video Content (T0180.004: Voyeuristic Content, T0170.003: Historic Content))), T0179.008: Content Solicits Action (T0162.001: Make Payment))))

Mumsnet targeted with child sexual abuse images

Field	Incident Properties
Name	Mumsnet targeted with child sexual abuse images
Live URL	https://www.bbc.co.uk/news/articles/c93qw3lw4kvo
Date Published (YYYY/MM/DD)	2025/02/04
Authors	David Mercer
Publication	BBC News
Archive URL	https://web.archive.org/web/20250204174848/https://www.bbc.com/news/articles/c93qw3lw4kvo
Summary	*Parenting site Mumsnet says it has stopped users from sharing pictures after it was targeted with images of child sexual abuse.*

Field	Incident Properties
Category	Documenting Sexual Assault, CSAM
Min Tags	T0180.006: Content Constitutes CSAM
Inline Tag	*Parenting site Mumsnet **(T0151.009: Legacy Online Forum Platform)** says it has stopped users from sharing pictures after it was targeted with images of child sexual abuse.* *Company founder Justine Roberts told the BBC the "horrific incident" had been reported to police after the images were posted on the platform over several hours late on Sunday.* *[...]* *"Several sets" of child abuse images **(T0156.001: Create Post (T0172.000: Image Content (T0180.006: Content Constitues CSAM)))** were posted on Mumsnet between 23:00 GMT on Sunday and 03:00 on Monday, Ms Roberts said.* *Most of the images were removed within an hour of being posted and all were taken down by 04:00 on Monday, she added **(T0151.009: Legacy Online Forum Platform (T0156.002: Delete Post))**.* *Ms Roberts said it was "pretty clear" there was an "ongoing, co-ordinated effort to further inflame the conversation on (the) site around the issue and to cause as much disruption as possible".*
Aggregate Text	An account posting CSAM to an online forum

Aggregate Brackets	(T0146.000: Account Asset (T0151.009: Legacy Online Forum Platform), T0156.001: Create Post (T0086: Image Content (T0180.006: Content Constitutes CSAM)))

‘It’s Disgusting’: Rosalía Fires Back at Artist Who Shared Photoshopped Nude Photos of Her

Field	Incident Properties
Name	‘It’s Disgusting’: Rosalía Fires Back at Artist Who Shared Photoshopped Nude Photos of Her
Live URL	https://www.rollingstone.com/music/music-news/rosalia-jc-reyes-photoshopped-pictures-1234740631/
Date Published (YYYY/MM/DD)	2023/05/24
Authors	Jon Blistein
Publication	Rolling Stone
Archive URL	https://web.archive.org/web/20230523223116/https://www.rollingstone.com/music/music-news/rosalia-jc-reyes-photoshopped-pictures-1234740631/
Summary	Rosalía vented her frustration with Spanish artist JC Reyes after he posted photoshopped images of her naked on social media, calling out the musician for not asking for consent and “creating a false narrative when I don’t even know you.”

Field	Incident Properties
Category	Individual Edited into Sexually Explicit Content
Min Tags	(T0176.007: Sexually Explicit Content, T0168.005: Third Party Introduced to Content)
Inline Tag	*[Spanish musician] Rosalía vented her frustration with Spanish artist JC Reyes after he posted photoshopped images of her naked on social media, calling out the musician for not asking for consent and “creating a false narrative when I don’t even know you.”* *[...]* *According to screengrabs circulating on social media, Reyes, or someone with access to his Instagram account, shared the photographs on his Stories. They appeared to be altered versions of photos Rosalía had originally taken and shared of herself* ***(T0156.001: Create Post (T0156.007: Time-Limited Post, T0172.000: Image Content (T0170.005: Content Produced by Third Party, T0170.007: Content Previously Published Online, T0168.004: Element Edited In to Content (T0176.007: Sexually Explicit Content))))*.** *[...]*

	While the photos have since been removed from Reyes' Instagram Stories, he seemed to boast about them — and suggest Rosalía had sent them to him ***(T0178.008: Fabricated Content Presented as Real, T0178.002: Content Presented as Produced by Third Party, T0178.001: Incorrect Content Source Presented)*** *— in a subsequent live video.*
Aggregate Text	A verified account on social media posting a time-limited post which contains an image previously posted online by a third party, which has been edited to introduce sexually explicit content
Aggregate Brackets	(T0146.003: Verified Account Asset (T0151.001: Social Media Platform), T0156.001: Create Post (T0156.007: Time-Limited Post, T0172.000: Image Content (T0170.005: Content Produced by Third Party, T0170.007: Content Previously Published Online, T0168.004: Element Edited In to Content (T0176.007: Sexually Explicit Content))))

'I was deepfaked by my best friend'

Field	Incident Properties
Name	'I was deepfaked by my best friend'
Live URL	https://www.bbc.com/news/uk-68673390
Date Published (YYYY/MM/DD)	2024/04/02
Authors	Kate West
Publication	BBC News
Archive URL	https://archive.ph/EVks6
Summary	*"Jodie" found images of herself used in deepfake porn - and then faced another terrible shock. She told BBC File on 4 about the moment she realised the person responsible was one of her best friends.*

Field	Incident Properties
Category	Sexually Explicit Deepfake Impersonation
Min Tags	(T0166.005: Deepfake Impersonation, T0176.007: Sexually Explicit Content)
Inline Tag	*In the spring of 2021, Jodie (not her real name) was sent a link to a porn website from an anonymous email account.* *Clicking through, she found explicit images and a video of what appeared to be her having sex with various men* ***(T0173.000: Video Content (T0166.005: Deepfake Impersonation, T0176.007: Sexually Explicit Content))****. Jodie's face* ***(T0166.004: Reference Content for AI-Generation)*** *had been digitally added onto*

	*another woman's body **(T0166.003: Source Content for AI-Generation)** - known as a "deepfake".* *Someone had posted photos of Jodie's face on a porn site saying she made them feel "so horny" and asking if other users on the site could make fake pornography of her **(T0156.001: Create Post (T0085: Text Content (T0179.008: Content Solicits Action (T0169.000: Produce Content (T0166.005: Deepfake Impersonation, T0176.007: Sexually Explicit Content))), T0172.000: Image Content (T0166.004: Reference Content for AI-Generation)))**. In exchange for the fakes, the user offered to share more photos of Jodie and details about her.* *Speaking for the first time about her experience, Jodie, who is now in her mid-20s, says, "I was screaming and crying and violently scrolling through my phone to work out what I was reading and what I was looking at."* *She adds: "I knew that this could genuinely ruin my life."* *Forcing herself to scroll through the porn site, Jodie said she felt her "whole world fall away".* *[...]* *It was not the first time Jodie had been targeted.* *In fact, it was the culmination of years of anonymous online abuse.* *When Jodie was a teenager, she discovered that her name and photos were being used on dating apps without her consent **(T0146: Account Asset (T0097.109: Romantic Suitor Persona (T0143.002: Fabricated Persona), T0143.003: Impersonated Persona, T0145.001: Copy Account Imagery, T0151.017: Dating Platform))**.* *This went on for years and she even received a Facebook message from a stranger in 2019 who said he was due to meet her at Liverpool Street station in London for a date.* *She told the man that it wasn't her who he had been speaking to. She says she felt "unnerved" because he knew all about who she was and had managed to find her online. He'd found her on Facebook after the "Jodie" on the dating app had stopped responding.* *In May 2020, during the UK's lockdown, Jodie was also alerted by a friend to a number of Twitter accounts that were posting pictures of her, with captions implying she was a sex worker.* *"What would you like to do with little teen Jodie?" read one caption next to an image of Jodie in a bikini, which had been taken from her private social media account **(T0146: Account Asset (T0151.008: Microblogging Platform), T0156.001: Create Post (T0085: Text Content (T0176.007: Sexually Explicit Content), T0172.000: Image Content (T0170.007: Content Previously Published Online, T0170.005: Content Produced by Third Party)))**.* *The Twitter handles posting these images had names like "slut exposer," and "chief perv." **(T0180.009: Discriminatory Content, T0180.008: Abusive Content)*** *All of the images being used were ones she'd been happy to share on her social media with close friends and family - but no one else.* *Then she found that these accounts were also posting images of other women she knew from university, as well as from her hometown of Cambridge.*

	"In that moment, I feel a very strong sense [that] I'm at the centre of this and this person is looking to hurt me," she said. *Jodie began to contact the other women in the pictures to warn them, including a close friend we are calling Daisy.* *"I just felt sick," said Daisy.* *Together the friends discovered many other Twitter accounts posting their images.* *"The more we looked, the worse it got," said Daisy.* *She messaged the Twitter users and asked where they had got their pictures. The reply was that the photos were "submissions" from anonymous senders who wanted them shared.* *"It's either an ex or someone who gets off on you," one user replied.* *Daisy and Jodie drew up a list of all the men who followed both of them on social media, and who could access both sets of their pictures.* *The friends concluded it must be Jodie's ex-boyfriend. Jodie confronted him and blocked him.* *For a few months, the posts stopped - but then an anonymous emailer got in touch.* *"Sorry to remain anonymous," the email read, "but I saw this guy was posting pics of you on creepy subreddits. I know this must be really scary."* *Jodie clicked on the link and was taken through to the online forum, Reddit, where a user had posted photos of Jodie and two of her friends, numbering them 1, 2 and 3.* *Others online were invited to take part in a game - which of these women would you have sex with, marry or kill.* *Beneath the post, 55 people had already commented.* *The photos used on the site were recent, and had been posted after Jodie blocked her ex. The women realised they had blamed the wrong person.* *Six weeks later, the same emailer got in touch again - this time about the deepfakes.* *When drawing up their list, Jodie and Daisy had ruled out a handful of men who they completely trusted, such as family - and Jodie's best friend, Alex Woolf.* *Jodie and Alex had struck up a firm friendship as teenagers, bonding over their shared love of classical music.* *Jodie had sought comfort from Woolf when she discovered that her name and photos were being used on dating apps without her consent.* *Woolf went on to get a double first in music from Cambridge University and won BBC Young Composer of the Year 2012, as well as appearing on Mastermind in 2021.*

	"He [Woolf] was very aware of the issues that faced women, especially on the internet," says Jodie. *"I really felt that he was an advocate."* *However, when she saw the deepfake porn photos, there was a picture of her in profile with the image of King's College, Cambridge, behind her* ***(T0166.003: Source Content for AI-Generation)***. *She clearly remembered it being taken - and that Woolf had also been in the photo. He was also the only other person she had shared the image with.* *It was Woolf who had been offering to share more original pictures of Jodie in exchange for them being turned into deepfakes.* *"He knew the impact that it was having on my life so profoundly," says Jodie. "And yet he still did it."* *In August 2021, Woolf, 26, was convicted of taking images of 15 women, including Jodie, from social media and uploading them to pornographic websites.* *He was given a 20-week prison sentence, suspended for two years and ordered to pay each of his victims £100 in compensation.* *Woolf has told the BBC he is "utterly ashamed" of the behaviour which led to his conviction and he is "deeply sorry" for his actions.* *"I think about the suffering I caused every day, and have no doubt that I will continue to do so for the rest of my life," he says.* *"There are no excuses for what I did, nor can I adequately explain why I acted on these impulses so despicably at that time."* *Woolf denies having anything to do with the harassment of Jodie which took place before the events he was charged with.* *For Jodie, finding out what her friend had done was the "ultimate betrayal and humiliation".* *She says: "I re-lived every conversation that we had, where he had comforted me and supported me and been kind to me. It was all a lie."* *We contacted X, formerly Twitter, and Reddit about the posts. X did not respond, but a spokesperson from Reddit said: "Non-consensual intimate media (NCIM) has no place on the Reddit platform* ***(T0175.006: Content Goes Against Platform Policy)***. *The subreddit in question has been banned* ***(T0151.010: Community Forum Platform, T0158.002: Delete Asset (T0151.011: Community Sub-Forum))***." *The porn site has also been taken down.* *In October 2023, sharing deepfake porn became a criminal offence as part of the Online Safety Bill.* *There are tens of thousands of deepfake videos online. Recent research found that 98% are pornographic, external.* *However, Jodie feels very angry that the new law does not criminalise a person who asks others to create*

	deepfakes, which is what Alex Woolf did. It is also not illegal to create a deepfake. *"This is affecting thousands of women and we need to have the proper laws and tools in place to stop people from doing this," she says.*
Aggregate Text	An account on a dating platform which is impersonating another individual, has copied their account imagery, and has fabricated a romantic suitor persona for them
Aggregate Brackets	(T0146: Account Asset (T0097.109: Romantic Suitor Persona (T0143.002: Fabricated Persona), T0143.003: Impersonated Persona, T0145.001: Copy Account Imagery, T0151.017: Dating Platform))
Aggregate Text	A post with text asking for the creation of a sexually explicit deepfake, along with an image to be used as a reference for the deepfake
Aggregate Brackets	(T0156.001: Create Post (T0085: Text Content (T0179.008: Content Solicits Action (T0169.000: Produce Content (T0166.005: Deepfake Impersonation, T0176.007: Sexually Explicit Content))), T0172.000: Image Content (T0166.004: Reference Content for AI-Generation)))

Woman's deepfake betrayal by close friend: 'Every moment turned into porn'

Field	Incident Properties
Name	Woman's deepfake betrayal by close friend: 'Every moment turned into porn'
Live URL	https://www.bbc.co.uk/news/articles/cm21j341m31o
Date Published (YYYY/MM/DD)	2025/02/08
Authors	Tiffanie Turnbull
Publication	BBC News
Archive URL	https://archive.ph/unhy2
Summary	*It was a warm February night when an ominous message popped into Hannah Grundy's inbox in Sydney.* *"I will just keep emailing because I think this is worthy of your attention," the anonymous sender wrote.* *Inside was a link, and a warning in bold: "[This] contains disturbing material."* *She hesitated for a moment, fearing it was a scam.* *The reality was so much worse. The link contained pages and pages of fake pornography featuring Hannah, alongside detailed rape fantasies and violent threats.*

	"You're tied up in them," she recalls. "You look afraid. You've got tears in your eyes. You're in a cage." *Written in kitschy word art on some images was Hannah's full name. Her Instagram handle was posted, as was the suburb she lived in. She would later learn her phone number had also been given out.* *That email kicked off a saga Hannah likens to a movie. She was left to become her own detective, uncovering a sickening betrayal by someone close to her, and building a case which changed her life - and Australian legal standards.*

Field	Incident Properties
Category	Sexually Explicit Deepfake Impersonation, Harassment
Min Tags	(T0166.005: Deepfake Impersonation, T0176.007: Sexually Explicit Content) (T0179.001: Content Threatens Action (T0127.003: Physical Sexual Violence, T0127.001: Physical Violence))
Inline Tag	*It was a warm February night when an ominous message popped into Hannah Grundy's inbox in Sydney.* *"I will just keep emailing because I think this is worthy of your attention," the anonymous sender wrote.* *Inside was a link, and a warning in bold: "[This] contains disturbing material."* ***(T0146.000: Account Asset (T0153.001: Email Platform), T0156.011: Send Message (T0085: Text Content (T0117.001: Link in Content)))*** *She hesitated for a moment, fearing it was a scam.* *The reality was so much worse. The link contained pages and pages of fake pornography featuring Hannah, alongside detailed rape fantasies and violent threats.* ***(T0152.004: Website Asset, T0163.005: Host Content (T0172.000: Image Content (T0166.005: Deepfake Impersonation, T0176.007: Sexually Explicit Content), T0085: Text Content (T0179.001: Content Threatens Action (T0127.003: Physical Sexual Violence, T0127.001: Physical Violence))))*** *"You're tied up in them," she recalls. "You look afraid. You've got tears in your eyes. You're in a cage."* *Written in kitschy word art on some images was Hannah's full name* ***(T0172.000: Image Content (T0166.005: Deepfake Impersonation, T0176.007: Sexually Explicit Content, T0085: Text Content (T0167.001: Personally Identifiable Information)))****. Her Instagram handle was posted, as was the suburb she lived in. She would later learn her phone number had also been given out .* *That email kicked off a saga Hannah likens to a movie. She was left to become her own detective, uncovering a sickening betrayal by someone close to her, and building a case which changed her life - and Australian legal standards.* *The web page was called "The Destruction of Hannah", and at the top of it was a poll where hundreds of people had voted on the vicious ways they wanted to abuse her* ***(T0029.000: Online Polls)****.* *Below was a thread of more than 600 vile photos, with Hannah's face stitched on to them. Buried in between*

	them were chilling threats. *"I'm closing in on this slut," the main poster said.* ***(T0180.009: Discriminatory Content)*** *"I want to hide in her house and wait until she is alone, grab her from behind and... feel her struggle."* ***(T0157.004: Comment on Post (T0179.001: Content Threatens Action (T0127.001: Physical Violence)))*** *It's been three years now, but the 35-year-old school teacher has no trouble recalling the "pure shock" that coursed through when she and partner Kris Ventura, 33, opened the page.* *"You immediately feel unsafe," Hannah tells me, eyes wide as she grips a mug of peppermint tea in her living room.* *Clicking through the website Kris had also found photos of their close friends, along with images depicting at least 60 other women, many also from Sydney.* *The couple quickly realised the pictures used to create the deepfakes* ***(T0172.000: Image Content (T0166.004: Reference Content for AI-Generation (T0170.007: Content Previously Published Online, T0170.005: Content Produced by Third Party)))*** *were from the women's private social media accounts. And the penny dropped: this was someone they all knew.*
Aggregate Text	An image containing sexually explicit deepfake impersonation, with text overlaid on the image which contains personally identifiable information
Aggregate Brackets	(T0172.000: Image Content (T0166.005: Deepfake Impersonation, T0176.007: Sexually Explicit Content, T0085: Text Content (T0167.001: Personally Identifiable Information)))

Sending Unsolicited Pictures of Your Penis Is a Form of Sexual Harassment. Please Stop Doing This

Field	Incident Properties
Name	Sending Unsolicited Pictures of Your Penis Is a Form of Sexual Harassment. Please Stop Doing This
Live URL	https://www.theroot.com/sending-unsolicited-pictures-of-your-penis-is-a-form-of-1802749386
Date Published (YYYY/MM/DD)	2017/09/08
Authors	Monique Judge
Publication	The Root
Archive URL	https://web.archive.org/web/20230712213321/https://www.theroot.com/sending-unsolicited-pictures-of-your-penis-is-a-form-of-1802749386

Summary	*The other day, I got a notification on Snapchat that a man I wasn't sure I knew added me. It said that he had added me by my user name, and his user name looked somewhat familiar, so, thinking it was possible we were already connected via social media, I added him back.* *Yesterday morning, when I woke up, I had a chat message from him. I clicked it, and to my horror, I was greeted with a video of his rather large penis being swung around like a weapon as some sort of advertisement for … I don't know what, exactly.* *I was livid. I did not know this person, and it was annoying that his means of introducing himself to me was sending an unsolicited video of his bare penis waving in the air like some sexual greeting card. I wasn't entertained. I wasn't amused. I wasn't intrigued, aroused or suddenly interested in him after having seen his penis. In fact, I felt attacked.*

Field	Incident Properties
Category	Unsolicited Sexual Imagery
Min Tags	(T0180.001: Unsolicited Sexual Imagery)
Inline Tag	*The other day, I got a notification on Snapchat that a man I wasn't sure I knew added me* ***(T0146.000: Account Asset (T0151.004: Chat Platform), T0161.001: Request Connection)****. It said that he had added me by my user name, and his user name looked somewhat familiar, so, thinking it was possible we were already connected via social media, I added him back* ***(T0146.000: Account Asset (T0151.004: Chat Platform), T0161.002: Connect with Account)****.* *Yesterday morning, when I woke up, I had a chat message from him. I clicked it, and to my horror, I was greeted with a video of his rather large penis being swung around like a weapon as some sort of advertisement for … I don't know what, exactly* ***(T0146.000: Account Asset (T0151.004: Chat Platform (T0153.007: Direct Messaging)), T0156.011: Send Message (T0173.000: Video Content (T0180.001: Unsolicited Sexual Imagery)))****.* *I was livid. I did not know this person, and it was annoying that his means of introducing himself to me was sending an unsolicited video of his bare penis waving in the air like some sexual greeting card. I wasn't entertained. I wasn't amused. I wasn't intrigued, aroused or suddenly interested in him after having seen his penis. In fact, I felt attacked.* *Unfortunately, this is not an uncommon occurrence for me. I am sex-positive, and I speak openly about sex and sex-related topics, so there are those who see this as an open invitation to throw pictures displaying their man parts either into my DMs or in my mentions on Twitter. I am always jarred when this happens.* *Think of social media as the streets of the internet. If you were to walk up to a woman in the streets of your city, randomly pull your penis out and wave it at her, you would be arrested and charged at the very least with indecent exposure. Why do you think it's OK to do this to women on the internet?* *Consent is an important issue that often gets overlooked because we live in a society steeped in rape culture. There is more victim-blaming than there is explaining to boys and men the proper ways to deal with their sexual urges. As I said recently on a podcast, men aren't always taught healthy ways to deal with their sexual urges and desires, and so they often act without impulse control.* *We women are tasked with dealing with these unwanted advances, skirting around fragile masculinities, and*

	being able to turn men down while fearing for our lives and our safety. This is a real thing, even on the internet. *Not too long ago, a man with a 14-inch penis plopped his gigantic monster in my mentions on Twitter* ***(T0146.000: Account Asset (T0151.008: Microblogging Platform (T0153.007: Direct Messaging)), T0156.011: Send Message (T0172.000: Image Content (T0180.001: Unsolicited Sexual Imagery)))****. I saw the picture, commented on the timeline (but not directly to him) that it looked scary to me and moved on about my business.* *About an hour later, he popped up in my mentions again, this time angrily asking me if I was even going to acknowledge it or say something to him* ***(T0146.000: Account Asset (T0151.008: Microblogging Platform (T0153.007: Direct Messaging)), T0156.011: Send Message (T0171.000: Text Content (T0179.008: Content Solicits Action (T0156.011: Send Message))))****. I asked him what he wanted me to say, and he told me he was upset that I had discussed his penis "like a science project" but not told him what I thought of it.* *I found this to be strange, and I told him that I had no interest in his penis whatsoever. His response to this was to send several more pictures and a video that showed him in action inside another woman* ***(T0146.000: Account Asset (T0151.008: Microblogging Platform (T0153.007: Direct Messaging)), T0156.011: Send Message (T0172.000: Image Content (T0180.001: Unsolicited Sexual Imagery), T0173.000: Video Content (T0180.001: Unsolicited Sexual Imagery))****.*
Aggregate Text	A message containing an unsolicited sexual image
Aggregate Brackets	(T0156.011: Send Message (T0172.000: Image Content (T0180.001: Unsolicited Sexual Imagery)))

Court jails first person convicted of cyberflashing in England

Field	Incident Properties
Name	Court jails first person convicted of cyberflashing in England
Live URL	https://www.theguardian.com/uk-news/2024/mar/19/court-jails-first-person-convicted-of-cyber-flashing-in-england
Date Published (YYYY/MM/DD)	2024/03/14
Authors	Jamie Grierson
Publication	The Guardian
Archive URL	https://web.archive.org/web/20240319132120/https://www.theguardian.com/uk-news/2024/mar/19/court-jails-first-person-convicted-of-cyber-flashing-in-england
Summary	*The first person in England to be convicted of a cyberflashing offence has been jailed for 66 weeks.*

	Nicholas Hawkes was convicted under the Online Safety Act after cyberflashing became an offence in England and Wales on 31 January. *The 39-year-old, from Basildon in Essex, was already a convicted sex offender when he sent unsolicited images of his genitals to a 15-year-old girl and a woman on 9 February, the Crown Prosecution Service said.*

Field	Incident Properties
Category	Unsolicited Sexual Imagery
Min Tags	(T0180.001: Unsolicited Sexual Imagery)
Inline Tag	*The first person in England to be convicted of a cyberflashing offence has been jailed for 66 weeks.* *Nicholas Hawkes was convicted under the Online Safety Act after cyberflashing became an offence in England and Wales on 31 January.* *The 39-year-old, from Basildon in Essex, was already a convicted sex offender when he sent unsolicited images of his genitals* ***(T0180.001: Unsolicited Sexual Imagery)*** *to a 15-year-old girl and a woman on 9 February, the Crown Prosecution Service said.* *Southend crown court heard on Tuesday that Hawkes asked to use his father's phone to call probation. He went into another room, where he sent the indecent photo via WhatsApp to a woman in her 60s* ***(T0146.000: Account Asset (T0151.004: Chat Platform), T0156.011: Send Message (T0172.000: Image Content (T0180.001: Unsolicited Sexual Imagery)))****. Minutes later, on the same device, he sent an explicit image to the child over iMessage* ***(T0146.000: Account Asset (T0151.004: Chat Platform), T0156.011: Send Message (T0172.000: Image Content (T0180.001: Unsolicited Sexual Imagery)))****, who was said to have been left "overwhelmed and crying".* *[...]* *Hawkes admitted during an earlier hearing at Southend magistrates' court to two counts of sending a photograph or film of genitals to cause alarm, distress or humiliation.* *Cyberflashing can involve offenders sending people an unsolicited sexual image on social media, dating apps, or via Bluetooth or Airdrop. Victims of the offence and other image-based abuses receive lifelong anonymity under the Sexual Offences Act.* *Hawkes was on the sex offender register after being convicted last year of sexual activity with a child under 16 and exposure, for which he received a community order.* *On Tuesday, he pleaded guilty to breaching the order and breaching a suspended sentence for another sexual offence.* *He was jailed for 66 weeks and handed a restraining order for the woman and the girl lasting 10 years, and a sexual harm prevention order banning him from approaching women who he does not know on public*

	highways and in parks for 15 years.
Aggregate Text	A message containing an unsolicited sexual image
Aggregate Brackets	(T0156.011: Send Message (T0172.000: Image Content (T0180.001: Unsolicited Sexual Imagery)))

My 'incel' attackers keep an online tally of their victims

Field	Incident Properties
Name	My 'incel' attackers keep an online tally of their victims
Live URL	https://www.bbc.co.uk/news/articles/c79dp383plwo
Date Published (YYYY/MM/DD)	2024/02/24
Authors	Anna O'Neill, Duc Ha
Publication	BBC News
Archive URL	https://archive.ph/TP9r7
Summary	*When Annie Makeeva set out from London on a solo trip to Vietnam in December 2022, she never imagined she would be sexually assaulted by a pair of violent "incels" - men who blame women because they are unable to find a sexual partner.*

Field	Incident Properties
Category	Documenting Sexual Assault
Min Tags	(T0180.011: Content Depicts Offline Harm (T0127.002: Physical Sexual Violence))
Inline Tag	*When Annie Makeeva set out from London on a solo trip to Vietnam in December 2022, she never imagined she would be sexually assaulted by a pair of violent "incels" - men who blame women because they are unable to find a sexual partner.* *It happened on the first day of her holiday, after Annie had cycled 10km (six miles) into the remote Cat Tien National Park, in the south of the country.* *"I decided to cycle along the jungle track and then hike to Crocodile Lake which is their star attraction. I stopped at the end of the track and I saw two guys there - fellow tourists, I assumed."* *But minutes later they followed her into the jungle and attacked her* ***(T0127.001: Physical Violence)****.*

	"As they walked past me the first man reached out and grabbed me. *"He then said something in Vietnamese to his friend who was on the other side of me. And I looked to see if this friend was coming to my rescue, or was he also going to attack me. And it turns out, yes, he wanted to attack me as well.* *"They were both groping me* ***(T0127.003: Physical Sexual Violence)****. I shouted for help and realised no-one could hear me."* *Annie, who is from west London, said she got "really scared" when the men began to restrain her and tried to push her down to the floor. She managed to break free and hide behind a tree.* *When one of the men again caught hold of her and looked back to his friend for a moment, she punched him hard on the back of the neck, disorientating him.* *At this point Annie got out her phone to take a photograph of the men, and this prompted the other man to offer her money to keep quiet.* *When she wouldn't take the money, the power balance shifted and the men began to back off. They made off the way they had come.* *Annie then had to decide whether to go in the same direction as the men to where she had left her bike, and cycle 10km back to the hotel, or try to go another 5km onwards to where she knew there would be a park ranger and potentially other tourists who might help her. She decided on the latter.* *[...]* *The men were eventually detained by police at the same hotel where Annie was staying.* *She says the police decided it was an opportunistic rather than planned attack and it is not clear what penalty, if any, they faced.* *[...]* *Struggling to come to terms with what had happened to her, Annie decided to try to find out whether the men had done this before.* *She had been encouraged to take a photo of the men's confiscated ID cards at the hotel and when she returned to the UK she used this information to look up their social media accounts.* *What she found horrified her.* *"It's full of violent imagery. It's extremely misogynistic, it promotes violence against women* ***(T0180.009: Discriminatory Content, T0179.002: Content Encourages Action (T0127.001: Physical Violence))****, violence against western tourists specifically.* *"There are pictures of weapons, including handmade weapons, target practice. They describe women as 'sluts' and 'livestock'* ***(T0180.009: Discriminatory Content, T0180.008: Abusive Content)****. And they also keep a tally of women they have attacked* ***(T0171.000: Text Content (T0180.011: Content Depicts Offline Harm (T0127.002: Physical Sexual Violence))****. They boast about their attacks and they make fun of us. It's*

	really chilling." *One of the men's social media accounts shows an image of a naked woman being beheaded and the words "the future of women".* ***(T0180.009: Discriminatory Content, T0179.001: Content Threatens Action (T0127.001: Physical Violence))*** *Another is full of obscene videos of him masturbating to pornography.* *"They're incels. They've posted on their social media accounts explaining that they haven't been able to find someone to date, someone to be intimate with. They have created this echo chamber of violence against women.* *"They made it very difficult for me to go back to my regular life. I stopped hiking. I didn't like to leave my house at night or really any time," says Annie, holding back tears.* *"I became hypervigilant and that is really exhausting. Can I cross the park to go to my yoga class? Who is behind me on the Tube? Can I even trust the colleague who gets in the lift with me?"* Images in the article show screenshots of the attackers' social media accounts and posts. One image shows an account on X owned by one of the attackers. Its cover photo is a cartoon showing two brown men in grass skirts carrying a pole which has a naked white woman tied to it. Its account description reads [translated from original in Vietnamese]: "like to go out, love animals, like naturally beautiful girls, have squeezed Western girls' breasts 8 times, local girls' breasts 3 times and Chinese girls' breasts 1 time". **(T0146.000: Account Asset (T0151.008: Microblogging Platform, T0145.005: Illustrated Character Account Imagery (T0180.011: Content Depicts Offline Harm (T0127.002: Physical Sexual Violence)))), T0158.001: Configure Asset Description (T0171.000: Text Content (T0180.011: Content Depicts Offline Harm (T0127.002: Physical Sexual Violence))))** Another image shows a post made to Instagram by an account owned by one of the attackers. The image appears to be a stock photo of a steamed bun. The text reads "Last night, it was raining white I was going out, according to the experience of thousands of lives of VIETNAMESE people, this is luck. Luckily, passing by the other street, saw the European tourist lady lining up a dumpling for her to bite. Took advantage of the busy road, got to touch a few things" **(T0146.000: Account Asset (T0151.001: Social Media Platform), T0156.001: Create Post (T0172.000: Image Content (T0170.004: Stock Media Content), T0085: Text Content (T0180.011: Content Depicts Offline Harm (T0127.002: Physical Sexual Violence), T0175.007: Coded Terminology)))**
Aggregate Text	Text which uses coded terminology to describe physical sexual violence
Aggregate Brackets	(T0085: Text Content (T0180.011: Content Depicts Offline Harm (T0127.002: Physical Sexual Violence), T0175.007: Coded Terminology))

Abuser first person jailed under Online Safety Act

Field	Incident Properties
Name	Abuser first person jailed under Online Safety Act
Live URL	https://www.bbc.co.uk/news/articles/cz91evw5kwzo
Date Published	2025/10/20

(YYYY/MM/DD)	
Authors	n/a
Publication	BBC News
Archive URL	https://web.archive.org/save/https://www.bbc.co.uk/news/articles/cz91evw5kwzo
Summary	A man who encouraged a schoolgirl he groomed and sexually abused to cut herself has become the first person to be jailed for encouraging self-harm under the Online Safety Act.
Why tag this report?	This report covers a man's use of online assets to groom a child, and sexually abuse her. The man was jailed under new laws for encouraging self-harm. Documenting assets and actions which were part of a conviction is useful for providing examples which lead to real consequences for threat actors - i.e. jail time.

Field	Incident Properties
Category	CSAM
Min Tags	(T0179.001: Content Threatens Action, T0179.008: Content Solicits Action (T0156.011: Send Message (T0172.000: Image Content (T0180.006: Content Constitutes CSAM))))
In-line tagging	Karl Davies, 42, a father-of-two from Wirral, pleaded guilty to and has been convicted of 17 offences relating to a school girl aged between 13 and 14. He showed "no pity and no mercy" for the "catastrophic" impact on his victim and had a "monstrous sense of sexual entitlement", Manchester Crown Court heard earlier. Davies, who has been jailed for 20 years, first made contact with the girl on Snapchat **(T0151.004: Chat Platform)** in June 2023, threatening and coercing her into sending him indecent images and videos of herself over the next year **(T0179.001: Content Threatens Action, T0179.008: Content Solicits Action (T0156.011: Send Message (T0172.000: Image Content (T0180.006: Content Constitutes CSAM))))**. Using several aliases to stay in contact, he then picked her up from school in "broad daylight" to sexually abuse her in his car on four occasions in the summer of 2024 **(T0127.003: Physical Sexual Violence)**. Sentencing Davies, Judge Hilary Manley imposed an extended sentence on Davies to reflect his "depraved and sadistic" crimes. During the year of abuse Davies also encouraged **(T0179.002: Content Encourages Action)** or coerced **(T0179.001: Content Threatens Action, T0179.008: Content Solicits Action)** his victim into filming herself committing acts of self-harm **(T0169.000: Produce Content (T0173.000: Video Content (T0127.001: Physical Violence)))** and bought her razors to facilitate this. [...] "Around June of 2023 a Snapchat account in the name of Ben Wild operated by the defendant, say the prosecution, added (the victim) in this case," **(T0146.000: Account Asset (T0151.004: Chat Platform, T0097.100: Individual Persona (T0143.002: Fabricated Persona)), T0161.001: Request Connection)** said Huw Edwards, prosecuting. Davies quickly ensured that the contact was "more than simply friendship" and sexual images were sent between the two, before he then put the girl in contact with another social media account.

	"The prosecution's case has always been that each of these accounts or aliases is in fact the defendant pretending to be a different person," said Mr Edwards. [...] Over time, the prosecution said, the aliases became "rather more sinister" threatening the girl **(T0179.001: Content Threatens Action)** and saying they would contact her father if she did not do as Davies asked. "Your offending displays a monstrous sense of sexual entitlement and a sinister desire for control," said Judge Manley. In June and July 2024, Davies drove to pick up the girl posing as "Mark" and made her perform sexual acts with him in four separate meetings **(T0127.003: Physical Sexual Violence)**. Mr Edwards said: "Before the meeting Joey - aka the defendant - told her what she would be doing with Mark. "Joey was the one who said to the victim you are to meet Mark and perform sexual acts upon him... She was told to record what they were doing, she did so and that was then shared with Joey on Snapchat." **(T0179.008: Content Solicits Action (T0169.000: Produce Content (T0173.000: Video Content (T0180.006: Content Constitutes CSAM))))** The judge added that Davies had used "trickery, lies, manipulations and blackmail" to satisfy his "twisted appetite" with a high level of control and planning.
Aggregate English	An account on a chat platform posing as an individual coerced a target over direct message to send them CSAM
Aggregate Brackets	(T0146.000: Account Asset (T0153.005: Direct Messaging (T0151.004: Chat Platform), T0097.100: Individual Persona (T0143.002: Fabricated Persona)), T0156.011: Send Message (T0179.001: Content Threatens Action, T0179.008: Content Solicits Action (T0156.011: Send Message (T0172.000: Image Content (T0180.006: Content Constitutes CSAM)))))

Annex 8 - DISARM TFGBV Playbook - Prefabricated Procedures

Technology-Facilitated Gender-Based Violence Playbook

Overview

Technology Facilitated Gender-Based Violence (TFGBV). Gender-based violence is "violence directed against a person because of that person's gender or violence that affects persons of a particular gender disproportionately." TFGBV is a "modern form of gender-based violence that utilizes digital technologies to cause harms".

Throughout this document the work of Suzie Dunn in her report Technology-Facilitated Gender-Based Violence - An Overview and the work of Jessica Ringrose, Kaitlyn Regehr and Betsy Milne in their report Understanding and Combatting Youth Experiences of Image-Based Sexual Harassment and Abuse will be quoted to give readers an introduction to different types of TFGBV.

Using this Document

This document has two sections; Core Behaviours and Detailed Tagging.

The Core Behaviours section contains templated tags which can be used to quickly capture common TFGBV behaviours. Analysts looking to broadly capture a harmful behaviour without going into specifics can apply tags identified here.

The Advanced Tagging section shows how analysts can modify tags to provide more detail about the behaviours they are seeing, and contains tagged examples of real-world reporting on TFGBV.

Core Behaviours

Behaviour	Tags
Non-consensual Distribution of Intimate Images	(T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery)
Voyeurism/Creepshots	(T00XX.0XX: Voyeuristic Content)
Sextortion	(T00XX.0XX: Content Constitutes Sextortion)
Documenting Sexual Assault	(T00XX.0XX: Produce Content (T00XX.0XX: Physical Sexual Violence))
Broadcasting Sexual Assault	(T00XX.0XX: Stream Content (T00XX.0XX: Physical Sexual Violence))
Unsolicited Sexual Imagery	(T00XX.0XX: Unsolicited Sexual Imagery)
AI Nudification	(T00XX.0XX: AI-Nudified Imagery)
Sexually Explicit Deepfake	(T00XX.0XX: Deepfake Impersonation, T00XX.0XX: Sexually Explicit Content)
Editing Individual into Sexual Imagery	(T0086: Image Content (T00XX.0XX: Sexually Explicit Content, T00XX.0XX: Third Party Introduced to Content))
Impersonation	(T0143.003: Impersonated Persona)
Threats	(T00XX.0XX: Content Threatens Action)
Hate Speech	(T00XX.0XX: Discriminatory Content, T00XX.0XX: Content Encourages Action (T00XX.0XX: Physical Violence))
Doxxing	(T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information)

Harassment	(T00XX.0XX: Abusive Content)
Networked Harassment	(T00XX.0XX: Post Content (T00XX.0XX: Abusive Content, T00XX.0XX: Networked Action))
Unsolicited Request for Sexual Imagery	(T00XX.0XX: Unsolicited Request for Sexual Imagery)
Unsolicited Comment on Appearance	(T00XX.0XX: Unsolicited Comment on Appearance)

This table contains templated tags which can be used to quickly capture common TFGBV behaviours.

Detailed Tagging

This section provides examples of more detailed tagging options for different types of TFGBV.

Image-Based Sexual Abuse

Non-consensual Distribution of Intimate Images

"The non-consensual distribution of intimate images ["Private Intimate Imagery"], which is often problematically described as revenge porn, occurs when a person's sexual images are shared with a wider than intended audience without the subject's consent" (Dunn 2020)

Behaviour	Tagging
Leaked image-based private intimate imagery	(T0086: Image Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery))
Leaked video-based private intimate imagery	(T0087: Video Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery))

A direct message which leaks image-based private intimate imagery	(T0153.007: Direct Messaging, T00XX.0XX: Send Message (T0086: Image Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery)))
A post which leaks personally identifiable information and image-based private intimate imagery	(T00XX.0XX: Create Post (T0085: Text Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information), T0086: Image Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery)))
A post which threatens the posting of private intimate imagery	(T00XX.0XX: Create Post (T0085: Text Content (T00XX.0XX: Content Threatens Action (T00XX.0XX: Create Post (T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery)))))
A website hosting image-based private intimate imagery	(T0152.004: Website Asset, T00XX.0XX: Host Content (T0086: Image Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery)))

Example: Telegram: Where women's nudes are shared without consent

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*In the split second Sara found out a nude photo of her had been leaked and shared on Telegram, her life changed. Her Instagram and Facebook profiles had been added, and her phone number included. Suddenly she was being contacted by unknown men asking for more pictures.*
Essential Tagging	(T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery, T00XX.0XX: Personally Identifiable Information) (T00XX.0XX: Unsolicited Request for Sexual Imagery)
Detailed Tagging: Inline	*In the split second Sara found out a nude photo of her had been leaked* ***(T0086: Image Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery))*** *and shared on Telegram, her life changed. Her Instagram and Facebook profiles had been added, and her*

	phone number included ***(T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information)****. Suddenly she was being contacted by unknown men asking for more pictures* ***(T00XX.0XX: Unsolicited Request for Sexual Imagery).***
Detailed Tagging: Summary	A person published a woman's private intimate imagery without her consent on Telegram alongside personally identifiable information (T0146: Account Asset (T0151.004: Chat Platform), T00XX.0XX: Send Message (T0085: Text Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information), T0086: Image Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery))). After this happened, she received messages asking her for sexual content (T00XX.0XX: Send Message (T00XX.0XX: Unsolicited Request for Sexual Imagery)).

[Back to top]

Voyeurism/Creepshots

"Voyeurism is defined as a person surreptitiously taking photos or recording a video of another person for a sexual purpose" (Dunn 2020). Also referred to as 'downblousing' or 'upskirting'.

Behaviour	Tagging
Image-based content produced for voyeuristic purposes	(T0086: Image Content (T00XX.0XX: Voyeuristic Content))
Video-based content produced for voyeuristic purposes	(T0087: Video Content (T00XX.0XX: Voyeuristic Content))
Non-consensually, secretly produced voyeuristic content depicting a consensual sex act	(T00XX.0XX: Voyeuristic Content, T00XX.0XX: Sexually Explicit Content)

A direct message containing a voyeuristic image, and a request to produce nudified imagery	(T0153.007: Direct Messaging, T00XX.0XX: Send Message (T0086: Image Content (T00XX.0XX: Voyeuristic Content), T0085: Text Content (T00XX.0XX: Content Solicits Action (T00XX.0XX: Produce Content (T00XX.0XX: AI-Nudified Imagery)))))
A website hosting video-based voyeuristic content	(T0152.004: Website Asset, T00XX.0XX: Host Content (T0087: Video Content (T00XX.0XX: Voyeuristic Content)))

Example: Upskirt photos shared in Facebook groups, BBC finds

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*This is a man following a schoolgirl in New York* *[Displayed here is a clip of footage produced by the actor; a vertical video recording taken from the point of view of a person following a child wearing a skirt up a set of stairs].* *He is about to film up her skirt and post the footage on Facebook.* *We found groups on Facebook where men share 'upskirt' photos.*
Essential Tagging	(T00XX.0XX: Voyeuristic Content, T00XX.0XX: Content Depicts Child Sexual Abuse)
Detailed Tagging: Inline	*This is a man following a schoolgirl in New York* *[Displayed here is a clip of footage produced by the actor; a vertical video recording taken from the point of view of a person following a child wearing a skirt up a set of stairs].* *He is about to film up her skirt* ***(T00XX: Produce Content (T0087.0XX: Vertical Video (T00XX.0XX: Voyeuristic Content, T00XX.0XX: Content Depicts Child Sexual Abuse)))*** *and post the footage on Facebook* ***(T0146: Account Asset (T0151.001: Social Media Platform), T00XX.0XX: Create Post (T0087.0XX: Vertical Video (T00XX.0XX: Voyeuristic Content, T00XX.0XX: Content Depicts Child Sexual Abuse)))****.*

	*We found groups on Facebook **(T0151.002: Online Community Group (T0151.001: Social Media Platform))** where men share 'upskirt' photos.*
Detailed Tagging: Summary	A man produced a video for his sexual pleasure by recording up the skirt of a child (T00XX: Produce Content (T0087.0XX: Vertical Video (T00XX.0XX: Voyeuristic Content, T00XX.0XX: Content Depicts Child Sexual Abuse))). He then posted the video to a Facebook group (T0146: Account Asset (T0151.001: Social Media Platform (T0151.002: Online Community Group)), T00XX.0XX: Create Post (T0087.0XX: Vertical Video (T00XX.0XX: Voyeuristic Content, T00XX.0XX: Content Depicts Child Sexual Abuse))).

[Back to top]

Sextortion

"Sexual extortion, or "sextortion," occurs when an individual has, or claims to have, a sexual image of another person and uses it to coerce a person into doing something they do not want to do" (Dunn 2020)

Detailed Tagging: Types of Sextortion

Sextortion is documented by combining a threat of non-consensually publishing sexual material depicting the target with a solicitation for the target to act.

Behaviour	Tagging
A threat of leaking private intimate imagery with a request for action	(T00XX.0XX: Content Constitutes Sextortion (T00XX.0XX: Content Threatens Action (T00XX.0XX: Create Post (T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery)), T00XX.0XX: Content Solicits Action))
A threat to post a sexually explicit deepfake with a request for action	(T00XX.0XX: Content Constitutes Sextortion (T00XX.0XX: Content Threatens Action (T00XX.0XX: Create Post (T00XX.0XX: Sexually Explicit Deepfake Impersonation)), T00XX.0XX: Content Solicits Action))

A threat to post sexually explicit voyeuristic content with a request for action	(T00XX.0XX: Content Constitutes Sextortion (T00XX.0XX: Content Threatens Action (T00XX.0XX: Create Post (T00XX.0XX: Voyeuristic Content, T00XX.0XX: Sexually Explicit Content)), T00XX.0XX: Content Solicits Action))

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging
A direct message which solicits payment with the threat of leaking intimate imagery	(T0153.007: Direct Messaging, T00XX.0XX: Send Message (T0085: Text Content (T00XX.0XX: Content Constitutes Sextortion (T00XX.0XX: Content Threatens Action (T00XX.0XX: Create Post (T0086: Image Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery))), T00XX.0XX: Content Solicits Action (T00XX.0XX: Make Payment))))).
A direct message which solicits access to a user's account with the threat of posting AI-Nudified Imagery	(T0153.007: Direct Messaging, T00XX.0XX: Send Message (T0085: Text Content (T00XX.0XX: Content Constitutes Sextortion (T00XX.0XX: Content Threatens Action (T00XX.0XX: Create Post (T0086: Image Content (T00XX.0XX: AI-Nudified Imagery))), T00XX.0XX: Content Solicits Action (T00XX.0XX: Send Message (T00XX.0XX: Asset Login Credentials)))))).

Example: Chinese actress Jiang Mengjie praised for revealing upskirting blackmail

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*Chinese actress Jiang Mengjie has been praised for sharing she was the victim of upskirting and was blackmailed over footage that circulated of her.* *Upskirting - involving a device such as a camera phone to take explicit images underneath a victim's clothing without permission - often goes undetected.* *Jiang Mengjie told her eight million followers on social network Weibo the video had been filmed "many years ago".* *[...]*

	Jiang further revealed that she had begun receiving private messages blackmailing her over the footage "saying that they would send the video to major film and TV companies and brands, and ruin the rest of my life". *[Here, a screenshot of messages is shown. In it, a user sends the messages "my patience is limited", "Sister, don't be so nervous. I'm just asking for money." "You have ignored me for so long, it seems you can afford the contract breach of contract, advertising breach of contract, and your acting career?]"*
Essential Tagging	(T00XX.0XX: Content Threatens Action (T00XX.0XX: Send Message (T00XX.0XX: Voyeuristic Content)), T00XX.0XX: Content Solicits Action (T00XX.0XX: Make Payment))
Detailed Tagging: Inline	*Chinese actress Jiang Mengjie has been praised for sharing she was the victim of upskirting and was blackmailed over footage that circulated of her.* *Upskirting - involving a device such as a camera phone to take explicit images underneath a victim's clothing without permission - often goes undetected.* *Jiang Mengjie told her eight million followers on social network Weibo the video had been filmed "many years ago"* ***(T0086: Video Content (T00XX.0XX: Voyeuristic Content, T00XX.0XX: Historic Content))****.* *[...]* *Jiang further revealed that she had begun receiving private messages* ***(T0153.007: Direct Messaging, T00XX.0XX: Send Message)*** *blackmailing her over the footage "saying that they would send the video to major film and TV companies and brands, and ruin the rest of my life"* ***(T00XX.0XX: Content Threatens Action (T00XX.0XX: Send Message (T0086: Video Content (T00XX.0XX: Voyeuristic Content, T00XX.0XX: Historic Content))))****.* *[Here, a screenshot of messages is shown. In it, a user sends the messages "my patience is limited", "Sister, don't be so nervous. I'm just asking for money."* ***(T00XX.0XX: Content Solicits Action (T00XX.0XX: Make Payment))*** *"You have ignored me for so long, it seems you can afford the contract breach of contract, advertising breach of contract, and your acting career?]"* ***(T00XX.0XX: Content Constitutes Sextortion)***
Detailed Tagging:	Threat actors previously produced and published voyeuristic materials (T00XX: Produce Content (T0086: Video Content (T00XX.0XX: Voyeuristic Content))) targeting the Chinese actress Jiang Mengjie.

Summary	Later, online actors threatened to send these images to organisations Mengjie works with in an attempt to coerce her into giving them money (T0153.007: Direct Messaging, T00XX.0XX: Send Message (T0085: Text Content (T00XX.0XX: Content Constitutes Sextortion (T00XX.0XX: Content Threatens Action (T00XX.0XX: Send Message (T0086: Video Content (T00XX.0XX: Voyeuristic Content, T00XX.0XX: Historic Content))), T00XX.0XX: Content Solicits Action (T00XX.0XX: Make Payment))))).

[Back to top]

Documenting Sexual Assault

"In cases of documenting or broadcasting sexual assault, the images of the assault are recorded and sometimes disseminated, resulting in an additional form of sexual violence against the victim-survivor" (Dunn 2020)

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging
Content which depicts sexual assault	(T00XX.0XX: Content Depicts Sexual Abuse)

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging
A video recording of a sexual assault	(T0087: Video Content (T00XX.0XX: Content Depicts Sexual Abuse))
Written description of a sexual assault	(T0085: Text Content (T00XX.0XX: Content Depicts Sexual Abuse))
A website hosting a video recording of a sexual assault	(T0152.004: Website Asset, T00XX.0XX: Host Content (T0087: Video Content (T00XX.0XX: Content Depicts Sexual Abuse)))

A direct message sending a video recording of a sexual assault	(T0153.007: Direct Messaging, T00XX.0XX: Send Message (T0087: Video Content (T00XX.0XX: Content Depicts Sexual Abuse)))

Example: My 'incel' attackers keep an online tally of their victims

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*When Annie Makeeva set out from London on a solo trip to Vietnam in December 2022, she never imagined she would be sexually assaulted by a pair of violent "incels" - men who blame women because they are unable to find a sexual partner.* *It happened on the first day of her holiday, after Annie had cycled 10km (six miles) into the remote Cat Tien National Park, in the south of the country.* *[...]* *Struggling to come to terms with what had happened to her, Annie decided to try to find out whether the men had done this before.* *She had been encouraged to take a photo of the men's confiscated ID cards at the hotel and when she returned to the UK she used this information to look up their social media accounts.* *What she found horrified her.* *"It's full of violent imagery. It's extremely misogynistic, it promotes violence against women, violence against western tourists specifically.* *"There are pictures of weapons, including handmade weapons, target practice. They describe women as 'sluts' and 'livestock'. And they also keep a tally of women they have attacked. They boast about their attacks and they make fun of us. It's really chilling."* *[...]* Images in the article show screenshots of the attackers' social media accounts and posts. One image shows an account on X owned by one of the attackers. Its cover photo

	is a cartoon showing two brown men in grass skirts carrying a pole which has a naked white woman tied to it. Its account description reads [translated from original in Vietnamese]: “like to go out, love animals, like naturally beautiful girls, have squeezed Western girls’ breasts 8 times, local girls’ breasts 3 times and Chinese girls’ breasts 1 time”.
Essential Tagging	(T0085: Text Content (T00XX.0XX: Content Depicts Sexual Abuse))
Detailed Tagging: Inline	*When Annie Makeeva set out from London on a solo trip to Vietnam in December 2022, she never imagined she would be sexually assaulted by a pair of violent "incels" - men who blame women because they are unable to find a sexual partner.* *It happened on the first day of her holiday, after Annie had cycled 10km (six miles) into the remote Cat Tien National Park, in the south of the country.* *[...]* *Struggling to come to terms with what had happened to her, Annie decided to try to find out whether the men had done this before.* *She had been encouraged to take a photo of the men's confiscated ID cards at the hotel and when she returned to the UK she used this information to look up their social media accounts.* *What she found horrified her.* *"It's full of violent imagery. It's extremely misogynistic, it promotes violence against women, violence against western tourists specifically.* *"There are pictures of weapons, including handmade weapons, target practice. They describe women as 'sluts' and 'livestock' **(T00XX.0XX: Discriminatory Content, T00XX.0XX: Abusive Content)**. And they also keep a tally of women they have attacked. They boast about their attacks and they make fun of us. **(T00XX.0XX: Content Depicts Sexual Abuse)** It's really chilling."* *[...]* Images in the article show screenshots of the attackers’ social media accounts and posts.

	One image shows an account on X owned by one of the attackers. Its cover photo is a cartoon showing two brown men in grass skirts carrying a pole which has a naked white woman tied to it. Its account description reads [translated from original in Vietnamese]: “like to go out, love animals, like naturally beautiful girls, have squeezed Western girls’ breasts 8 times, local girls’ breasts 3 times and Chinese girls’ breasts 1 time”. **(T0146: Account Asset ((T0151.008: Microblogging Platform)(T0145.005: Illustrated Character Account Imagery (T00XX.0XX: Content Depicts Sexual Abuse))), T00XX.0XX: Configure Asset Description (T00XX.0XX: Content Depicts Sexual Abuse))**
Detailed Tagging: Summary	Two men sexually assaulted a woman (T00XX.0XX: Physical Sexual Violence). One of the men kept a tally of sexual assaults he had committed in his account’s biography field (T0146: Account Asset (T0151.008: Microblogging Platform), T00XX.0XX: Configure Asset Description (T00XX.0XX: Content Depicts Sexual Abuse))

Back to top

Broadcasting Sexual Assault

“In cases of documenting or broadcasting sexual assault, the images of the assault are recorded and sometimes disseminated, resulting in an additional form of sexual violence against the victim-survivor” (Dunn 2020)

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging
A livestream of a sexual assault	(T00XX.0XX: Stream Content (T00XX.0XX: Content Depicts Sexual Abuse))

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging

An account on a social media platform livestreaming a sexual assault	T0146: Account Asset (T0151.001: Social Media Platform), T00XX.0XX: Stream Content (T0087: Video Content (T00XX.0XX: Content Depicts Sexual Abuse))
An account on a social media platform streaming historic footage of a sexual assault	T0146: Account Asset (T0151.001: Social Media Platform), T00XX.0XX: Stream Content (T0087: Video Content (T00XX.0XX: Content Depicts Sexual Abuse, T00XX.0XX: Historic Content))

Example: Facebook Live 'broadcasts gang rape' of woman in Sweden

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*Three men were arrested on suspicion of rape in Sweden on Sunday, following reports of an assault against a woman being live-streamed on Facebook.* *Police in Uppsala were contacted in the morning by a woman who said she had seen a gang rape broadcast in a closed group on the site.* *"You have been raped," one of the men said at the end of the video and then laughed, according to the viewer.* *Police later confirmed they, and "many" others, had seen the footage.* *The Facebook group is said to have several thousand members.* *Police confirmed that they had found three men, aged between 19 and 25, and one woman at a local apartment.* *The men were arrested on the spot.*
Essential Tagging	(T00XX.0XX: Stream Content (T00XX.0XX: Content Depicts Sexual Abuse))
Detailed Tagging: Inline	*Three men were arrested on suspicion of rape in Sweden on Sunday, following reports of an assault against a woman being live-streamed on Facebook.* *Police in Uppsala were contacted in the morning by a woman who said she had seen a gang rape broadcast* ***(T0146: Account Asset (T0151.001: Social Media Platform), T00XX.0XX: Stream Content (T0087: Video Content (T00XX.0XX: Content Depicts Sexual Abuse)))*** *in a closed group* ***(T0151.002: Online Community Group (T0155: Gated Asset))*** *on the site.*

	"You have been raped," one of the men said at the end of the video and then laughed, according to the viewer. *Police later confirmed they, and "many" others, had seen the footage.* *The Facebook group is said to have several thousand members.* *Police confirmed that they had found three men, aged between 19 and 25, and one woman at a local apartment.* *The men were arrested on the spot.*
Detailed Tagging: Summary	Three men raped a woman (T00XX.0XX: Physical Sexual Violence), and used an account on Facebook to livestream the assault to a closed Facebook group (T0146: Account Asset (T0151.001: Social Media Platform (T0151.002: Online Community Group (T0155: Gated Asset))), T00XX.0XX: Stream Content (T0087: Video Content (T00XX.0XX: Content Depicts Sexual Abuse))).

[Back to top]

Unsolicited Sexual Imagery

"Also referred to as 'cyberflashing' (McGlynn & Johnson, 2020), the sending of unwanted sexual images most often involves adult cisgendered men sending unsolicited images of their genitals (i.e. 'unsolicited dick pics') over digital technologies, such as Air Drop, social media platforms, dating platforms or video conferencing platforms" (Ringrose, Regehr, Milne 2021)

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging
Unsolicited Sexual Imagery	(T00XX.0XX: Unsolicited Sexual Imagery)

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging
Sending an unsolicited sexual image over direct message	(T0153.007: Direct Messaging, T00XX.0XX: Send Message (T0086: Image Content (T00XX.0XX: Unsolicited Sexual Imagery)))
An account on Snapchat sending a single-view unsolicited sexual image	(T0146: Account Asset (T0146: Chat Platform (T0153.007: Direct Messaging)), T00XX.0XX: Send Message (T00XX.0XX: View-Limited Post)(T0086: Image Content (T00XX.0XX: Unsolicited Sexual Imagery))

Example: Court jails first person convicted of cyberflashing in England

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*Nicholas Hawkes was convicted under the Online Safety Act after cyberflashing became an offence in England and Wales on 31 January.* *The 39-year-old, from Basildon in Essex, was already a convicted sex offender when he sent unsolicited images of his genitals to a 15-year-old girl and a woman on 9 February, the Crown Prosecution Service said.* *Southend crown court heard on Tuesday that Hawkes asked to use his father's phone to call probation. He went into another room, where he sent the indecent photo via WhatsApp to a woman in her 60s. Minutes later, on the same device, he sent an explicit image to the child over iMessage, who was said to have been left "overwhelmed and crying".*
Essential Tagging	(T00XX.0XX: Unsolicited Sexual Imagery)
Detailed Tagging: Inline	*Nicholas Hawkes was convicted under the Online Safety Act after cyberflashing became an offence in England and Wales on 31 January.* *The 39-year-old, from Basildon in Essex, was already a convicted sex offender when he sent unsolicited images of his genitals* ***(T00XX.0XX: Unsolicited Sexual Imagery)*** *to a 15-year-old girl and a woman on 9 February, the Crown Prosecution Service said.* *Southend crown court heard on Tuesday that Hawkes asked to use his father's phone to call probation. He went into another room, where he sent the indecent*

	photo via WhatsApp to a woman in her 60s ***(T0146: Account Asset (T0151.004: Chat Platform), T00XX.0XX: Send Message (T0086: Image Content (T00XX.0XX: Unsolicited Sexual Imagery)))****. Minutes later, on the same device, he sent an explicit image to the child over iMessage* ***(T0146: Account Asset (T0151.004: Chat Platform), T00XX.0XX: Send Message (T0086: Image Content (T00XX.0XX: Unsolicited Sexual Imagery)))****, who was said to have been left "overwhelmed and crying".*
Detailed Tagging: Summary	A man sent unsolicited photos of his genitals to a woman and a child using WhatsApp and iMessage (T0146: Account Asset (T0151.004: Chat Platform), T00XX.0XX: Send Message (T0086: Image Content (T00XX.0XX: Unsolicited Sexual Imagery))).

[Back to top]

Synthetic Media

AI Nudification

"Nudify" apps allow users to upload an image of a person (a "source image"), which is then edited by the platform using AI to produce a version of the image in which the person is naked.

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging
AI-Nudified media	(T00XX.0XX: AI-Nudified Imagery)

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging

An AI-Nudified Image	(T0086: Image Content (T00XX.0XX: AI-Nudified Imagery))
Base media submitted to an AI for nudification	(T00XX.0XX: Source Content for AI-Generation)
An account on an AI-Media platform using the platform to create a nudified image	(T0146: Account Asset (T0154.002: AI-Media Platform), T00XX: Produce Content (T0086: Image Content (T00XX.0XX: AI-Nudified Imagery)))
An account on an AI-Media platform uploading an image to be nudified	(T0146: Account Asset (T0154.002: AI-Media Platform), T00XX.0XX: Upload File (T0086: Image Content (T00XX.0XX: Source Content for AI-Generation)))
An account on an AI-Media platform uploading an image, and using the platform to nudify the image	(T0146: Account Asset (T0154.002: AI-Media Platform), T00XX.0XX: Upload File (T0086: Image Content (T00XX.0XX: Source Content for AI-Generation)), T00XX: Produce Content (T0086: Image Content (T00XX.0XX: AI-Nudified Imagery)))
A member of a chat group requesting members to produce a nudified version of a submitted image, alongside the target's personally identifiable information	(T0146: Account Asset (T0151.004: Chat Platform (T0151.006: Chat Room)), T00XX.0XX: Send Message (T0086: Image Content, T0085: Text Content)
A website hosting a nudified image	(T0152.004: Website Asset, T00XX.0XX: Host Content (T0086: Image Content (T00XX.0XX: AI-Nudified Imagery)))

Example: Ads on Instagram and Facebook for a deepfake app undressed a picture of 16-year-old Jenna Ortega

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*Facebook and Instagram hosted ads that featured a blurred fake nude image of an underage celebrity used to promote an app that billed itself as a way to make sexually explicit images with artificial intelligence.* *A review of Meta's ad library showed that the company behind the app ran 11 ads that used a manipulated, blurred photo of "Wednesday" actor Jenna Ortega, taken when she was 16 years old. The ads appeared on the two platforms as well as its Messenger app for most of February. The app, called Perky AI, advertised that it could undress women with artificial intelligence.*

	The ads showed how the Perky app could change Ortega's outfit in the photo based on text prompts, including "Latex costume," "Batman underwear" and finally, "No clothes."
Essential Tagging	(T00XX.0XX: AI-Nudified Imagery, T00XX.0XX: Content Depicts Child Sexual Abuse)
Detailed Tagging: Inline	*Facebook and Instagram hosted ads that featured a blurred fake nude image of an underage celebrity **(T0151.001: Social Media Platform, T0114: Deliver Ad (T0086: Image Content (T00XX.0XX: AI-Nudified Imagery, T00XX.0XX: Content Depicts Child Sexual Abuse)))** used to promote an app that billed itself as a way to make sexually explicit images with artificial intelligence.* *A review of Meta's ad library showed that the company behind the app ran 11 ads that used a manipulated, blurred photo of "Wednesday" actor Jenna Ortega, taken when she was 16 years old **(T00XX.0XX: Source Content for AI-Generation)**. The ads appeared on the two platforms as well as its Messenger app **(T0151.004: Chat Platform)** for most of February. The app, called Perky AI **(T0154.002: AI Media Platform)**, advertised that it could undress women with artificial intelligence.* *The ads showed how the Perky app could change Ortega's outfit in the photo based on text prompts **(T00XX.0XX: Prompt for AI-Generation)**, including "Latex costume," "Batman underwear" and finally, "No clothes."*
Detailed Tagging: Summary	Facebook and Instagram delivered ads which included a nudified image of a 16-year old girl (T0151.001: Social Media Platform, T0114: Deliver Ad (T0086: Image Content (T00XX.0XX: AI-Nudified Imagery, T00XX.0XX: Content Depicts Child Sexual Abuse))).

[Back to top](#)

Sexually Explicit Deepfake Impersonation

"Employing artificial intelligence, one can swap a person's face onto the face of another person in a [sexually explicit video], making it appear as though they are featured in the sexual video performing sex acts they never participated in" (Dunn 2020)

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging
A sexually explicit deepfake impersonation	(T00XX.0XX: Sexually Explicit Deepfake Impersonation)

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging
Image-based sexually explicit deepfake	(T0086: Image Content (T00XX.0XX: Sexually Explicit Deepfake Impersonation))
Video-based sexually explicit deepfake	(T0087: Video Content (T00XX.0XX: Sexually Explicit Deepfake Impersonation))
Media depicting the target of impersonation provided to an AI for reference in creating a deepfake	(T00XX.0XX: Reference Content for AI-Generation)
Media depicting sexually explicit scenes to provided to an AI to edit the impersonated target into	(T00XX.0XX: Source Content for AI-Generation, T00XX.0XX: Sexually Explicit Content)
A text prompt soliciting a sexually explicit deepfake impersonation submitted to an AI-Platform, alongside images of the target	(T0146: Account Asset (T0154.002: AI Media Platform), T00XX: Send Message (T0085: Text Content (T00XX.0XX: Prompt for AI-Generation, T00XX.0XX: Content Solicits Action (T00XX: Produce Content (T00XX.0XX: Sexually Explicit Deepfake Impersonation))))(T0086: Image Content (T00XX.0XX: Reference Content for AI-Generation)))
A post containing a sexually explicit deepfake impersonation presenting the content as genuine	(T00XX.0XX: Create Post (T0086: Image Content (T00XX.0XX: Sexually Explicit Deepfake Impersonation), T0085: Text Content (T00XX.0XX: Fabricated Content Presented as Real)))

A reply to a post containing a sexually explicit deepfake impersonation, and text degrading the target of the impersonation	(T00XX.0XX: Comment on Post (T0086: Image Content (T00XX.0XX: Sexually Explicit Deepfake Impersonation), T0085: Text Content (T00XX.0XX: Abusive Content)))
A website hosting an image-based sexually explicit deepfake impersonation	(T0152.004: Website Asset, T00XX.0XX: Host Content (T0086: Image Content (T00XX.0XX: Sexually Explicit Deepfake Impersonation)))

Example: Deepfake Creators Are Revictimizing GirlsDoPorn Sex Trafficking Survivors

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*Over the past two months, an account on the largest deepfake sexual abuse website has posted 12 celebrity videos that are based on footage from GirlsDoPorn, a now-defunct sex trafficking operation that the US Department of Justice says its operators used to conspire and commit sex trafficking through "force, fraud, and coercion," tricking five women—and allegedly hundreds more—into making sex videos that were subsequently posted online.* *The dozen videos—which ran up to 21 minutes long and racked up tens of thousands of views before they were taken down following WIRED's inquiry—used footage originally posted to the GirlsDoPorn website and had celebrity faces added using artificial intelligence.*
Essential Tagging	(T00XX.0XX: Sexually Explicit Deepfake Impersonation, T00XX.0XX: Content Depicts Sexual Abuse)
Detailed Tagging: Inline	*Over the past two months, an account on the largest deepfake sexual abuse website has posted 12 celebrity videos* ***(T00XX.0XX: Sexually Explicit Deepfake Impersonation)*** *that are based on footage from GirlsDoPorn, a now-defunct sex trafficking operation that the US Department of Justice says its operators used to conspire and commit sex trafficking through "force, fraud, and coercion,"* ***(T00XX.0XX: Physical Sexual Violence)*** *tricking five women—and allegedly hundreds more— into making sex videos* ***(T00XX: Produce Content (T0087: Video Content (T00XX.0XX: Content Depicts Sexual Abuse)))*** *that were subsequently posted online* ***(T00XX.0XX: Create Post (T0087: Video Content (T00XX.0XX: Content Depicts Sexual Abuse)))****.* *The dozen videos—which ran up to 21 minutes long and racked up tens of thousands of views before they were taken down following WIRED's*

	*inquiry—used footage originally posted to the GirlsDoPorn website **(T00XX.0XX: Source Content for AI-Generation)** and had celebrity faces added using artificial intelligence **(T00XX.0XX: Reference Content for AI-Generation)**.*
Detailed Tagging: Summary	An account on a website hosting sexually explicit deepfakes posted 12 video-based sexually explicit deepfake impersonations (T0146: Account Asset (T0152.004: Website Asset), T00XX.0XX: Create Post (T0087: Video Content (T00XX.0XX: Sexually Explicit Deepfake Impersonation))) The videos were produced using footage of sexual assaults as source content (T0087: Video Content (T00XX.0XX: Content Depicts Sexual Abuse, T00XX.0XX: Source Content for AI-Generation)).

[Back to top]

Individual Edited into Sexually Explicit Imagery

"Early examples of the misuse of technology to create synthetic sexual images include utilizing Photoshop to superimpose a person's face on the body of a sexual image (Delfino 2019), the practice of which remains fairly common in Bangladesh, India and Pakistan" (Dunn 2020)

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging
A sexually explicit image	(T0086: Image Content (T00XX.0XX: Sexually Explicit Content))
An image edited to introduce a third party	(T0086: Image Content (T00XX.0XX: Third Party Introduced to Content))
A sexually explicit image edited to introduce a third party	(T0086: Image Content (T00XX.0XX: Sexually Explicit Content, T00XX.0XX: Third Party Introduced to Content))
A non-sexual image edited to be sexually explicit	(T0086: Image Content (T00XX.0XX: Element Edited In to Content (T00XX.0XX: Sexually Explicit Content)))

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging
A post containing an edited sexual image presenting the content as genuine	(T00XX.0XX: Create Post (T0086: Image Content (T00XX.0XX: Sexually Explicit Content, T00XX.0XX: Third Party Introduced to Content), T0085: Text Content (T00XX.0XX: Fabricated Content Presented as Real)))
A reply to a post containing an edited sexual image, and text degrading the target of the impersonation	(T00XX.0XX: Comment on Post (T0086: Image Content (T00XX.0XX: Sexually Explicit Content, T00XX.0XX: Third Party Introduced to Content), T0085: Text Content (T00XX.0XX: Abusive Content)))
A website hosting an edited sexual image	(T0152.004: Website Asset, T00XX.0XX: Host Content (T0086: Image Content (T00XX.0XX: Sexually Explicit Content, T00XX.0XX: Third Party Introduced to Content)))

Example: 'It's Disgusting': Rosalía Fires Back at Artist Who Shared Photoshopped Nude Photos of Her

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*[Spanish musician] Rosalía vented her frustration with Spanish artist JC Reyes after he posted photoshopped images of her naked on social media, calling out the musician for not asking for consent and "creating a false narrative when I don't even know you."* *[...]* *According to screengrabs circulating on social media, Reyes, or someone with access to his Instagram account, shared the photographs on his Stories. They appeared to be altered versions of photos Rosalía had originally taken and shared of herself.* *[...]* *While the photos have since been removed from Reyes' Instagram Stories, he seemed to boast about them — and suggest Rosalía had sent them to him — in a subsequent live video.*

Essential Tagging	(T0086: Image Content (T00XX: Element Edited In to Content (T00XX.0XX: Sexually Explicit Content)))
Detailed Tagging: Inline	*[Spanish musician] Rosalía vented her frustration with Spanish artist JC Reyes after he posted photoshopped images of her naked on social media, calling out the musician for not asking for consent and "creating a false narrative when I don't even know you."* *[...]* *According to screengrabs circulating on social media, Reyes, or someone with access to his Instagram account, shared the photographs on his Stories. They appeared to be altered versions of photos Rosalía had originally taken and shared of herself* ***(T0146.003: Verified Account Asset (T0151.001: Social Media Platform), T00XX.0XX: Create Post (T00XX.0XX: Time-Limited Post, T0086: Image Content (T00XX.0XX: Content Produced by Third Party, T00XX.0XX: Content Previously Published Online, T00XX: Element Edited In to Content (T00XX.0XX: Sexually Explicit Content))))*. *[...]* *While the photos have since been removed from Reyes' Instagram Stories, he seemed to boast about them — and suggest Rosalía had sent them to him* ***(T00XX.0XX: Fabricated Content Presented as Real, T00XX.0XX: Content Presented as Produced by Third Party, T00XX.0XX: Incorrect Content Source Presented)*** *— in a subsequent live video.*
Detailed Tagging: Summary	A man posted took images a woman posted to her social media, edited the images to make the woman appear naked, and posted them to his Instagram Stories (T0146.003: Verified Account Asset (T0151.001: Social Media Platform), T00XX.0XX: Create Post (T00XX.0XX: Time-Limited Post, T0086: Image Content (T00XX.0XX: Content Produced by Third Party, T00XX.0XX: Content Previously Published Online, T00XX: Element Edited In to Content (T00XX.0XX: Sexually Explicit Content)))). During a later livestream he claimed that the images were real, and had been sent to him by the targeted woman (T00XX.0XX: Stream Content (T0087: Video Content (T00XX.0XX: Fabricated Content Presented as Real, T00XX.0XX: Content Presented as Produced by Third Party, T00XX.0XX: Incorrect Content Source Presented))).

[Back to top]

Impersonation

"Impersonation can lead to reputational damage and put a person at physical risk. Some abusers have created fake online accounts of women to spread false information and damage the reputation of the person they are impersonating (Gurumurthy, Vasudevan and Chami 2019). Abusers have created fake websites impersonating the victim-survivor in an attempt to ruin their personal relationships and destroy their job prospects (Dunn, forthcoming 2021). They may also send fake messages from the victim-survivor's accounts or fake accounts to damage their personal and professional relationships (Freed et al. 2017). A study from Bangladesh, India and Pakistan found that women who had lower incomes or were younger or sexual minorities were more likely to be impersonated and that the impersonation often had a sexual element (Sambasivan et al. 2019)." (Dunn 2020)

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging
Impersonation	(T0143.003: Impersonated Persona)
An account falsely presenting itself as being controlled by a targeted person	(T0146: Account Asset (T0143.003: Impersonated Persona))

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging
An account impersonating another person, and falsely presenting that person as a romantic suitor	(T0146: Account Asset ((T0143.003: Impersonated Persona)(T0097.109: Romantic Suitor Persona (T0143.002: Fabricated Persona))))
An account on a dating platform impersonating an individual, sending messages that solicit	(T0146: Account Asset ((T0151.017: Dating Platform)(T0143.003: Impersonated Persona)(T0097.109: Romantic Suitor Persona (T0143.002: Fabricated Persona))), T00XX.0XX: Send

in-person sexual meetups at the impersonated target's address	Message (T0085: Text Content ((T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information)(T00XX.0XX: Content Solicits Action (T00XX.0XX: Offline Gathering, T00XX.0XX: Sexually Explicit Content)))))

Example: Stalkers use online sex ads as weapon

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*The 33-year-old mother of four had divorced Michael Anthony Johnson II, an unemployed computer specialist from Hyattsville, in 2011. Their relationship was tempestuous.* *Johnson went to her home one night in 2011, got in her car and waited for her until the next morning. When she got in, Johnson wrapped his hands around her neck. She escaped, and Johnson was convicted of assault.* *Now, she believed Johnson had her in his sights again. When she logged onto Craigslist in the days after the man showed up at her home, she found ad after ad. They had increasingly vile titles including one that read: "Rape Me and My Daughters."* *When she clicked on the ad, her photo popped up and her address was listed.* *[...]* *In March, the woman reported the harassment to the Library of Congress, which began investigating its employee. An agent with the Office of the Inspector General responded to one of the ads.* *"Can we meet up today?" the agent asked, according to court records.* *"Sure can here's my address" the poster wrote back. "Just a side note my gate has been giving fits you may have to park and walk up my lane, sorry." It was signed "Love" and the woman's name. A photo of the woman was attached.* *Investigators later determined that the message was sent from an IP address at the Library of Congress's facility in Culpeper where Kuban worked and that his e-mail was used to post ads on Craigslist, according to court records.* *[...]*

	The day after she discovered the "Rape me" listing, another stomach-churning ad was posted. It offered her then-12- and 13-year-old daughters and 12-year-old son up for sex in exchange for cash. The children's photos appeared in the ad. *But the digital assault was just beginning. The woman found fake profiles for herself on a host of sites including Facebook and the pornography aggregator XTube, soliciting men for sex and listing her address.*
Essential Tagging	(T0143.003: Impersonated Persona) (T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information) (T00XX.0XX: Content Solicits Action (T00XX.0XX: Physical Sexual Violence))
Detailed Tagging: Inline	*The 33-year-old mother of four had divorced Michael Anthony Johnson II, an unemployed computer specialist from Hyattsville, in 2011. Their relationship was tempestuous.* *Johnson went to her home one night in 2011, got in her car and waited for her until the next morning. When she got in, Johnson wrapped his hands around her neck* ***(T00XX.0XX: Physical Violence)****. She escaped, and Johnson was convicted of assault.* *Now, she believed Johnson had her in his sights again. When she logged onto Craigslist in the days after the man showed up at her home, she found ad after ad. They had increasingly vile titles including one that read: "Rape Me and My Daughters."* *When she clicked on the ad, her photo popped up and her address was listed* ***(T0146: Account Asset ((T0152.004: Website Asset)(T0143.003: Impersonated Persona)(T0097.109: Romantic Suitor Persona (T0143.002: Fabricated Persona))), T00XX.0XX: Create Post (T0085: Text Content ((T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information), T00XX.0XX: Content Solicits Action (T00XX.0XX: Physical Sexual Violence)), T0086: Image Content))****.* *[...]* *In March, the woman reported the harassment to the Library of Congress, which began investigating its employee. An agent with the Office of the Inspector*

	General responded to one of the ads. *"Can we meet up today?" the agent asked, according to court records.* *"Sure can here's my address" the poster wrote back. "Just a side note my gate has been giving fits you may have to park and walk up my lane, sorry." It was signed "Love" and the woman's name. A photo of the woman was attached.* ***(T0146: Account Asset ((T0152.004: Website Asset)(T0143.003: Impersonated Persona)(T0097.109: Romantic Suitor Persona (T0143.002: Fabricated Persona))), T00XX.0XX: Send Message (T0085: Text Content ((T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information), T00XX.0XX: Content Solicits Action (T00XX.0XX: Offline Gathering)), T0086: Image Content))****.* *Investigators later determined that the message was sent from an IP address at the Library of Congress's facility in Culpeper* ***(T0149.006: IP Address Asset)*** *where Kuban worked and that his e-mail* ***(T0146: Account Asset (T0153.001: Email Platform))*** *was used to post ads on Craigslist, according to court records.* *[...]* *The day after she discovered the "Rape me" listing, another stomach-churning ad was posted. It offered her then-12- and 13-year-old daughters and 12-year-old son up for sex in exchange for cash. The children's photos appeared in the ad.* ***(T0146: Account Asset ((T0152.004: Website Asset)(T0143.003: Impersonated Persona)), T00XX.0XX: Create Post (T0085: Text Content (T00XX.0XX: Content Solicits Action (T00XX.0XX: Physical Sexual Violence, T00XX.0XX: Make Payment)), T0086: Image Content))****.* *But the digital assault was just beginning. The woman found fake profiles for herself on a host of sites including Facebook* ***(T0146: Account Asset ((T0143.003: Impersonated Persona)(T0151.001: Social Media Platform)))*** *and the pornography aggregator XTube* ***(T0146: Account Asset ((T0143.003: Impersonated Persona)(T0152.006: Video Platform (T00XX.0XX: Adult Entertainment Outlet Persona, T0143.001: Authentic Persona))))****, soliciting men for sex and listing her address* ***((T00XX.0XX: Content Solicits Action (T00XX.0XX: Offline Gathering, T00XX.0XX: Sexually Explicit Content)) (T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information))****.*
Detailed	A man impersonated a woman online and created posts which solicited sexual assault, and included the woman's address (T0146: Account Asset ((T0152.004:

Tagging: Summary	Website Asset)(T0143.003: Impersonated Persona)(T0097.109: Romantic Suitor Persona (T0143.002: Fabricated Persona))), T00XX.0XX: Create Post (T0085: Text Content ((T00XX.0XX: Leak of Private Material. T00XX.0XX: Personally Identifiable Information), T00XX.0XX: Content Solicits Action (T00XX.0XX: Physical Sexual Violence)), T0086: Image Content)).

[Back to top]

Threats

"Death threats and rape threats have become common and even normalized in online dialogue (Van der Wilk 2018). Research by Safety Net Canada (2013) found that threats and intimidation were the most commonly reported forms of TFGBV against victim services workers in Canada" (Dunn 2020)

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging
A threat	(T00XX.0XX: Content Threatens Action)

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging
Threat of violence	(T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Violence))
Threat of sexual violence	(T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Sexual Violence))
Threat of leaking sensitive material	(T00XX.0XX: Content Threatens Action (T00XX.0XX: Create Post (T00XX.0XX: Leak of Private Material)))
A reply to a post which threatens sexual violence	(T00XX.0XX: Comment on Post (T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Sexual Violence)))

A direct message which threatens sexual violence, and includes the target's physical address	(T0153.007: Direct Messaging, T00XX.0XX: Send Message (T0085: Text Content (T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Sexual Violence), T00XX.0XX: Personally Identifiable Information)))

Example: I still won't ignore internet rape and death threats - Lauren Mayberry

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*Lauren Mayberry was 27 when she was sent rape threats after wearing a black mini dress in a music video.* *The frontwoman of synth pop band Chvrches had become the target of a torrent of online abuse, following the release of the track Leave a Trace from the band's acclaimed second album.* *One person offered to hand her a gun, if she couldn't deal with the backlash.* *[...]* *In 2019 she received rape and death threats after expressing disappointment in EDM star Marshmello's decision to work with Chris Brown and Tyga.* *Brown, who previously admitted assaulting his then girlfriend Rihanna, hit back at the band, calling them a "bunch of losers".* *Tyga, who was sued for sexual assault, said: "Everyone makes mistakes".* *Mayberry shared some of the messages she received following the exchange on social media.* *One person said he would "bash her skull in" if she continued to make comments about Brown*
Essential Tagging	(T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Violence, T00XX.0XX: Physical Sexual Violence))
Detailed Tagging: Inline	*Lauren Mayberry was 27 when she was sent rape threats* ***(T00XX.0XX: Send Message (T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Sexual Violence)))*** *after wearing a black mini dress in a music video.*

	The frontwoman of synth pop band Chvrches had become the target of a torrent of online abuse, following the release of the track Leave a Trace from the band's acclaimed second album. *One person offered to hand her a gun, if she couldn't deal with the backlash* ***(T00XX.0XX: Send Message (T00XX.0XX: Content Encourages Action (T00XX.0XX: Physical Violence)))***. *[...]* *In 2019 she received rape and death threats* ***(T00XX.0XX: Send Message (T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Violence, T00XX.0XX: Physical Sexual Violence)))*** *after expressing disappointment in EDM star Marshmello's decision to work with Chris Brown and Tyga.* *Brown, who previously admitted assaulting his then girlfriend Rihanna, hit back at the band, calling them a "bunch of losers".* *Tyga, who was sued for sexual assault, said: "Everyone makes mistakes".* *Mayberry shared some of the messages she received following the exchange on social media.* *One person said he would "bash her skull in" if she continued to make comments about Brown* ***(T00XX.0XX: Send Message (T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Violence)))***.
Detailed Tagging: Summary	A woman was targeted with threats of sexual violence after publishing a music video in which she wore a black mini dress (T00XX.0XX: Send Message (T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Sexual Violence))). Later, the same woman was targeted with threats of physical and sexual violence after she voiced her opinion on collaborating with a man who had physically assaulted his girlfriend (T00XX.0XX: Send Message (T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Violence, T00XX.0XX: Physical Sexual Violence))).

[Back to top](#)

Hate Speech

"Hate speech is a particularly abhorrent form of TFGBV that dehumanizes and encourages violence toward a person or a group of people based on an identifying feature, such as their religion, gender, ethnicity, disability or other identity factor" (Dunn 2020)

Incident: Technology-Facilitated Gender-Based Violence - An Overview
Incident: My 'incel' attackers keep an online tally of their victims

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging
Content which discriminates against individuals or groups based on characteristics such as race, gender, religion, or sexual orientation	(T00XX.0XX: Discriminatory Content)
Content which encourages violence	(T00XX.0XX: Content Encourages Action (T00XX.0XX: Physical Violence))
Abusive content (i.e. unwanted communication which causes mental distress or fear to the recipient)	(T00XX.0XX: Abusive Content)
Hate speech (i.e. discriminatory content that encourages violence towards a group)	(T00XX.0XX: Discriminatory Content, T00XX.0XX: Content Encourages Action (T00XX.0XX: Physical Violence))

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging
A comment on a post claiming that women who demand equal rights as men should be slapped	(T0146: Account Asset, T00XX.0XX: Comment on Post (T0085: Text Content (T00XX.0XX: Discriminatory Content, T00XX.0XX: Content Encourages Action (T00XX.0XX: Physical Violence))))

A direct message claiming women who demand equal rights as men need to be sexually assaulted	(T0146: Account Asset (T0153.007: Direct Messaging), T00XX.0XX: Send Message (T0085: Text Content (T00XX.0XX: Discriminatory Content, T00XX.0XX: Content Encourages Action (T00XX.0XX: Physical Sexual Violence))))

Example: "Female stupidity at its best. They all need to die.": Violent and sexualised hate speech targeting women approved for publication by social media platforms

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*Together with independent public interest law centre in South Africa, the Legal Resources Centre, we carried out a joint investigation looking at Facebook, TikTok, X/Twitter, and YouTube's ability to detect and remove real-world examples of hate speech targeting women journalists.* *Rather than publishing the examples on the platforms as user content, we submitted them to all four platforms in the form of adverts, so they could be scheduled in the future and so that we could remove them before going live.* *This methodology is designed to test the platforms' first line of defence against hate speech, before advertising is published, giving us an indication of their ability to identify and moderate actual hate speech that is live on the platform.* *The test consisted of 10 adverts in four languages: English, Afrikaans, Xhosa, and Zulu (40 adverts total). Real-world examples of misogynistic hate speech were edited to clarify language and grammar, none were coded or difficult to interpret, text was illustrated by video footage, and all clearly violated the platforms' advertising policies.* *The content followed the platforms' own definitions of hate speech outlined in their policies: all targeted women specifically and were violent, dehumanising, expressed inferiority, contempt, and disgust. For example, the adverts referred to women as prostitutes, psychopaths, or vermin, and called for them to be beaten and killed.* *Nearly all the test adverts were approved for publication by all four platforms. Meta and TikTok approved all 40 ads within 24 hours.*
Essential Tagging	(T00XX.0XX: Abusive Content, T00XX.0XX: Discriminatory Content, T00XX.0XX: Content Encourages Action (T00XX.0XX: Physical Violence))

Detailed Tagging: Inline	*Together with independent public interest law centre in South Africa, the Legal Resources Centre, we carried out a joint investigation looking at Facebook* ***(T0151.001: Social Media Platform)****, TikTok, X/Twitter* ***(T0151.008: Microblogging Platform)****, and YouTube's* ***(T0152.006: Video Platform)*** *ability to detect and remove real-world examples of hate speech targeting women journalists.* *Rather than publishing the examples on the platforms as user content, we submitted them to all four platforms in the form of adverts* ***(T0114: Deliver Ad)****, so they could be scheduled in the future and so that we could remove them before going live.* *This methodology is designed to test the platforms' first line of defence against hate speech, before advertising is published, giving us an indication of their ability to identify and moderate actual hate speech that is live on the platform.* *The test consisted of 10 adverts in four languages* ***(T0101: Create Localised Content)****: English, Afrikaans, Xhosa, and Zulu (40 adverts total). Real-world examples of misogynistic hate speech were edited to clarify language and grammar, none were coded or difficult to interpret, text was illustrated by video footage, and all clearly violated the platforms' advertising policies* ***(T00XX.0XX: Content Goes Against Platform Policy)****.* *The content followed the platforms' own definitions of hate speech outlined in their policies: all targeted women specifically and were violent, dehumanising, expressed inferiority, contempt, and disgust. For example, the adverts referred to women as prostitutes, psychopaths, or vermin* ***(T00XX.0XX: Abusive Content, T00XX.0XX: Discriminatory Content)****, and called for them to be beaten and killed* ***(T00XX.0XX: Content Encourages Action (T00XX.0XX: Physical Violence))****.* *Nearly all the test adverts were approved for publication by all four platforms. Meta and TikTok approved all 40 ads within 24 hours.*
Detailed Tagging: Summary	Facebook (T0151.001: Social Media Platform), TikTok, X/Twitter (T0151.008: Microblogging Platform), and YouTube (T0152.006: Video Platform) reviewed and approved ads which went against their stated platform policy against hate speech (T0146: Deliver Ad (T0085: Text Content (T00XX.0XX: Content Goes Against Platform Policy, T00XX.0XX: Abusive Content, T00XX.0XX: Discriminatory Content, T00XX.0XX: Content Encourages Action (T00XX.0XX: Physical Violence)), T0087: Video Content))

[Back to top]

Doxing

"Doxing is the publication of personal information such as a person's legal name, address, phone number, contact information, driver's licence, workplace, and private documents or correspondence without their consent." (Dunn 2020). Also referred to as 'doxxing', or 'doxxed'.

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging
Personal information such as a person's legal name, address, phone number, contact information, driver's licence, workplace,	(T00XX.0XX: Personally Identifiable Information)
Private Documents	(T00XX.0XX: Internal Documents)
Private Correspondence	(T00XX.0XX: Internal Communications)
Account login credentials	(T00XX.0XX: Asset Login Credentials)
Private materials which have been published without the owner's consent	(T00XX.0XX: Leak of Private Material)
Leaked personally identifiable information	(T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information)

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging
A post which doxes a target alongside a call to harass them	(T00XX.0XX: Create Post (T0085: Text Content ((T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information) (T00XX.0XX: Content Solicits Action (T00XX.0XX: Send Message (T00XX.0XX: Abusive Content))))))

A post which doxes a target alongside a leak of their private intimate imagery	(T00XX.0XX: Create Post (T0085: Text Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information)), T0086: Image Content (T00XX.0XX: Private Intimate Imagery, T00XX.0XX: Leak of Private Material))

Example: Rana Ayyub, the face of India's women journalists plagued by cyber-harassment

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*When a video depicting a stabbing of a bishop in Wakeley, Australia was uploaded to X and Meta-owned platforms in April 2024, Inman Grant issued a takedown notice to the platforms. These powers are derived from Australian law, which prohibits content depicting "acts of terrorism," and allows her to subsequently request certain illegal content be removed from online platforms. Meta complied with the request within the hour, Inman Grant said, but X kept the content up, despite the fact that it likely violated the platform's violent content policy. When the Federal Court of Australia granted an interim injunction compelling X Corp to hide the violent material, Musk began tweeting about Inman Grant on April 22, 2024, calling her an "unelected official" and "eSafety Commissar," evoking authoritarian sentiments and claiming that Inman Grant "demand[ed] *global* content bans[.]" These dog whistles—"the use of words or symbols with a double (or coded) meaning that is abusive or harmful, sometimes to signal a group of online abusers to attack a specific target" —to his 192 million followers led to increased, targeted harassment against Inman Grant.* *[...]* *Inman Grant's experience reflects the offline impact of enemy images. Users threatened her, her family, and her employees. OpenAI's sentiment analysis model assessed a full 10 percent of content in this dataset as threatening. For example, one user wrote: "@tweetinjules Vile, white-hating, racist pos. You are one fugly man. Thankfully you marxists will soon be wiped out." In this environment of hate, Inman Grant's family members were doxxed and users directed credible death threats at her, necessitating the involvement of the Australian Federal Police.*
Essential Tagging	(T00XX.0XX: Leak of Private Materials, T00XX.0XX: Personally Identifiable Information) (T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Violence)) (T00XX.0XX: Content Remains Active after Reporting Breach of Platform Policy to

	Platform) (T00XX.0XX: Unsolicited Comment on Appearance, T00XX.0XX: Abusive Content, T00XX.0XX: Discriminatory Content)
Detailed Tagging: Inline	*When a video depicting a stabbing of a bishop in Wakeley, Australia was uploaded to X and Meta-owned platforms in April 2024, Inman Grant issued a takedown notice to the platforms* ***(T0097.111: Government Official Persona (T0143.001: Authentic Persona)), T00XX.0XX: Send Message (T00XX.0XX: Content Solicits Action (T00XX.0XX: Delete Post)))****. These powers are derived from Australian law, which prohibits content depicting "acts of terrorism," and allows her to subsequently request certain illegal content be removed from online platforms. Meta complied with the request within the hour* ***(T0151: Social Media Platform, T00XX.0XX: Delete Post)****, Inman Grant said, but X kept the content up, despite the fact that it likely violated the platform's violent content policy* ***(T0151.008: Microblogging Platform, T00XX.0XX: Content Remains Active after Reporting Breach of Platform Policy to Platform)****. When the Federal Court of Australia granted an interim injunction compelling X Corp to hide the violent material, Musk began tweeting about Inman Grant on April 22, 2024, calling her an "unelected official" and "eSafety Commissar," evoking authoritarian sentiments and claiming that Inman Grant "demand[ed] *global* content bans[.]" These dog whistles—"the use of words or symbols with a double (or coded) meaning that is abusive or harmful, sometimes to signal a group of online abusers to attack a specific target" —to his 192 million followers led to increased, targeted harassment against Inman Grant.* *On April 23, 2024, there were 73,694 total mentions of Inman Grant or the eSafety Commissioner's office on X. By comparison, the office and Commissioner's average daily mentions on X for April through December 2023 were 145.* *In order to better understand this harassment and its gendered aspects, the research team used Meltwater, a social listening and sentiment analysis tool, to download a dataset encompassing posts that mentioned "Julie Inman Grant" or Inman Grant's X handle, "@tweetinjules,"* ***(T00XX.0XX: Tag Account)*** *from April 22-23.* *[...]* *Inman Grant's experience reflects the offline impact of enemy images. Users threatened her, her family, and her employees. OpenAI's sentiment analysis model assessed a full 10 percent of content in this dataset as threatening* ***(T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Violence))****. For example, one*

	user wrote: "@tweetinjules Vile, white-hating, racist pos. You are one fugly man. Thankfully you marxists will soon be wiped out." ***(T0146: Account Asset (T0151.008: Microblogging Platform), T00XX.0XX: Create Post (T0085: Text Content (T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Violence), (T00XX.0XX: Unsolicited Comment on Appearance, T00XX.0XX: Abusive Content, T00XX.0XX: Discriminatory Content), T00XX.0XX: Tag Account)))*** *In this environment of hate, Inman Grant's family members were doxxed* ***(T00XX.0XX: Leak of Private Materials, T00XX.0XX: Personally Identifiable Information)*** *and users directed credible death threats at her* ***(T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Violence))****, necessitating the involvement of the Australian Federal Police.*
Detailed Tagging: Summary	After Elon Musk tweeted about Inman Grant, there was a sharp spike in material referencing her on X. Posts doxxed her family members (T0146: Account Asset (T0151.008: Microblogging Platform), T00XX.0XX: Create Post (T0085: Text Content (T00XX.0XX: Leak of Private Materials, T00XX.0XX: Personally Identifiable Information))), and included threats of physical violence (T0146: Account Asset (T0151.008: Microblogging Platform), T00XX.0XX: Create Post (T0085: Text Content (T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Violence)))).

[Back to top]

Harassment

"Harassment encompasses a variety of unwanted digital communication (Duggan 2017; Digital Rights Foundation 2018). It can involve a brief incident, such as a single targeted racist or sexist comment (Lenhart et al. 2016), or a long-term organized attack, such as the Gamergate campaign [...]

"In its 2016 report on online harassment, Data & Society stated that "online harassment is defined less by the specific behavior than its intended effect on and the way it is experienced by its target" (Lenhart et al. 2016). Online harassment is known to cause the recipient mental distress and sometimes fear (Citron 2014)." (Dunn 2020)

Incident: 'His Facebook was a shrine to my face': the day I caught my catfish
Incident: 'I was deepfaked by my best friend'

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging
Abusive content (i.e. unwanted communication which causes mental distress or fear to the recipient)	(T00XX.0XX: Abusive Content)

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging
Abusive content which makes reference to characteristics such as race, gender, religion, or sexual orientation	(T00XX.0XX: Abusive Content, T00XX.0XX: Discriminatory Content)
Abusive content which makes reference to a target's physical appearance	(T00XX.0XX: Abusive Content, T00XX.0XX: Unsolicited Comment on Appearance)
An abusive comment left on a post	(T0146: Account Asset, T00XX.0XX: Comment on Post (T00XX.0XX: Abusive Content))

Example: Women's football subculture of misogyny: the escalation to online gender-based violence

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	This research examines comments left on videos showing female football players posted to TikTok by two UK-football teams' official accounts; Burnley Football Club and Manchester United. This section of the report looks at comments under the theme of 'misogyny and hatred of women': *Misogyny consists of animosity against women, as well as physical, psychological, and social aggression towards them (Code, 2002), and can be defined as hatred or disdain towards women (Moloney & Love, 2018). Misogynistic comments were apparent in all of the TikTok posts containing*

female football players. These differed from general sexist comments in their more aggressive tone and reliance on outdated gender stereotypes, which appeared to provoke anger when these stereotypes of women's traditional role were challenged. For example:

> *'Like we want a signed woman shirt. Your (sic) dogwater in football'*
>
> *'Why the fuck would you buy that shite' (referring to a women's shirt)*
>
> *'Who the fucking hell are you?'*
>
> *'Seems a bit desperate'*
>
> *'You're shit at football'*

As shown, these comments featured much more anger and rage at women's football being professionalised. These comments are meant to be very threatening to the female players, creating a toxic environment. This supports extant literature that recognises online gender-based violence (Jane, 2017, 2020; Powell & Henry, 2019; Richardson-Self, 2018).

The analysis also revealed examples of misogynistic comments that harked back to traditional gendered roles included:

> *'Do my dishes'*
>
> *'When did woman (sic) start playing footy? I was wondering why the ironing was building up'*
>
> *'We asked for varane not a cook'*
>
> *'Wooman ☕'*
>
> *'Women ☕hahahaha'*

This concurs with Fielding-Lloyd and Woodhouse (Fielding-Lloyd & Woodhouse, 2023), who argue that social media is used to perpetuate sexist football stereotypes. Our study builds on this by providing further insights into the types of sexism and misogyny occurring on football clubs' official TikTok accounts. The immersion in this environment revealed that the men who made these comments felt threatened by the changing

	nature of women's roles and their seeming equality to men in a traditionally male-dominated sport. *It is noted that many of the misogynistic comments were coupled with statements that attempted to present these interactions as humour. The cups of tea emojis refer to outdated sexist sayings about how women should make cups of tea for men and in the workplace. Similarly, the dishes comments refer to old-fashioned stereotypes of women staying in the home and doing the housework whilst men go out to work. Violent interactions mixed with the inclusion of 'jokes' online are recognised to act as a mask to the severity of the interaction (Kavanagh et al., 2019). Lockyer and Savigny (2020) suggest it is an important part of online discourse to recognise the adoption of humour as a tool adopted in order to normalise and/or trivialise gender-based discrimination, while Cole (2015) suggests that comments worded with humour are implemented to neutralise the sense of threat. These views are echoed in our findings in this study.*
Essential Tagging	(T00XX.0XX: Abusive Content, T00XX.0XX: Discriminatory Content)
Detailed Tagging: Inline	This research examines comments left on videos showing female football players posted to TikTok by two UK-football teams' official accounts; Burnley Football Club and Manchester United. This section of the report looks at comments under the theme of 'misogyny and hatred of women': *Misogyny consists of animosity against women, as well as physical, psychological, and social aggression towards them (Code, 2002), and can be defined as hatred or disdain towards women (Moloney & Love, 2018). Misogynistic comments were apparent in all of the TikTok posts containing female football players. These differed from general sexist comments in their more aggressive tone and reliance on outdated gender stereotypes, which appeared to provoke anger when these stereotypes of women's traditional role were challenged* ***(T0146: Account Asset (T0151.008: Microblogging Platform), T00XX.0XX: Comment on Post (T0085: Text Content (T00XX.0XX: Abusive Content, T00XX.0XX: Discriminatory Content)))****. For example:* *'Like we want a signed woman shirt. Your (sic) dogwater in football'* *'Why the fuck would you buy that shite' (referring to a women's shirt)*

'Who the fucking hell are you?'

'Seems a bit desperate'

'You're shit at football'

As shown, these comments featured much more anger and rage at women's football being professionalised. These comments are meant to be very threatening to the female players, creating a toxic environment. This supports extant literature that recognises online gender-based violence (Jane, 2017, 2020; Powell & Henry, 2019; Richardson-Self, 2018).

The analysis also revealed examples of misogynistic comments that harked back to traditional gendered roles included:

'Do my dishes'

'When did woman (sic) start playing footy? I was wondering why the ironing was building up'

'We asked for varane not a cook'

'Wooman 

'Women  *hahahaha'*

This concurs with Fielding-Lloyd and Woodhouse (Fielding-Lloyd & Woodhouse, 2023), who argue that social media is used to perpetuate sexist football stereotypes. Our study builds on this by providing further insights into the types of sexism and misogyny occurring on football clubs' official TikTok accounts. The immersion in this environment revealed that the men who made these comments felt threatened by the changing nature of women's roles and their seeming equality to men in a traditionally male-dominated sport.

It is noted that many of the misogynistic comments were coupled with statements that attempted to present these interactions as humour. The cups of tea emojis refer to outdated sexist sayings about how women should make cups of tea for men and in the workplace. Similarly, the dishes comments refer to old-fashioned stereotypes of women staying in the home and doing the housework whilst men go out to work. Violent

	interactions mixed with the inclusion of 'jokes' online are recognised to act as a mask to the severity of the interaction (Kavanagh et al., 2019). Lockyer and Savigny (2020) suggest it is an important part of online discourse to recognise the adoption of humour as a tool adopted in order to normalise and/or trivialise gender-based discrimination, while Cole (2015) suggests that comments worded with humour are implemented to neutralise the sense of threat. These views are echoed in our findings in this study.
Detailed Tagging: Summary	Accounts on TikTok managed by UK Football clubs posted videos depicting their female football teams (T0146.003: Verified Account Asset (T0151.008: Microblogging Platform), T00XX.0XX: Create Post (T0087.0XX: Vertical Video)). People left abusive, misogynistic comments on these videos (T0146: Account Asset (T0151.008: Microblogging Platform), T00XX.0XX: Comment on Post (T0085: Text Content (T00XX.0XX: Abusive Content, T00XX.0XX: Discriminatory Content))).

[Back to top]

Networked Harassment

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging
Action taken by multiple accounts which are coordinating	(T00XX: Networked Action)
Action taken within a short timeframe by multiple accounts which are coordinating	(T00XX.0XX: Concurrent Networked Action)

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging

Coordinated commenting on posts with harassment	(T0146: Account Asset, T00XX.0XX: Comment on Post ((T00XX: Networked Action)(T0085: Text Content (T00XX.0XX: Abusive Content))))
Coordinated direct messaging which threatens physical violence	(T0146: Account Asset (T0153.007: Direct Messaging), T00XX.0XX: Send Message ((T00XX: Networked Action)(T0085: Text Content (T00XX.0XX: Content Threatens Action (T00XX.0XX: Physical Violence)))))

Example: Rana Ayyub, the face of India's women journalists plagued by cyber-harassment

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*On the night of 8 November [2024], Rana Ayyub received over "200 calls and obscene messages," as she told RSF, just a few minutes after @HPhobiaWatch, an account on the social media platform X, published Rana Ayyub's personal telephone number, encouraging his followers to harass her. The @HPhobiaWatch account is managed by an anonymous influencer affiliated with Hindutva, a Hindu supremacist movement, posting under the pseudonym Hindutva Knight,* *Following this initial wave of attacks, a screenshot of a pornographic deepfake of Rana Ayyub was circulated on social networks. Soon after, other accounts linked to the same nationalist movement leaked the journalist's identity documents and the passwords to her social media accounts. False rumours and tweets attributing false statements to her went viral.*
Essential Tagging	(T00XX.0XX: Send Message ((T00XX.0XX: Concurrent Networked Action)(T00XX.0XX: Abusive Content))) (T00XX.0XX: Leak of Private Material, T00XX.0XX: Internal Documents, T00XX.0XX: Asset Login Credentials, T00XX.0XX: Personally Identifiable Information) (T00XX.0XX: Content Encourages Action (T00XX.0XX: Send Message (T00XX.0XX: Abusive Content)))
Detailed Tagging: Inline	*On the night of 8 November [2024], Rana Ayyub received over "200 calls and obscene messages," **(T00XX.0XX: Send Message ((T00XX.0XX: Concurrent Networked Action)(T00XX.0XX: Abusive Content)))** as she told RSF, just a few minutes after @HPhobiaWatch, an account on the social media platform X, published Rana Ayyub's personal telephone number, encouraging his followers to harass her **(T0146: Account Asset (T0151.001: Social Media Platform), T00XX.0XX: Create Post ((T00XX.0XX: Leak of Private Material, T00XX.0XX:***

	Personally Identifiable Information), T00XX.0XX: Content Encourages Action (T00XX.0XX: Send Message (T00XX.0XX: Abusive Content))))*. The @HPhobiaWatch account is managed by an anonymous influencer affiliated with Hindutva, a Hindu supremacist movement, posting under the pseudonym Hindutva Knight,* *Following this initial wave of attacks, a screenshot of a pornographic deepfake of Rana Ayyub was circulated on social networks* ***(T00XX.0XX: Create Post (T0086.0XX: Screenshot (T00XX.0XX: Sexually Explicit Deepfake Impersonation)))****. Soon after, other accounts linked to the same nationalist movement leaked the journalist's identity documents and the passwords to her social media accounts* ***(T00XX.0XX: Create Post ((T00XX: Networked Action)(T00XX.0XX: Leak of Private Material, T00XX.0XX: Internal Documents, T00XX.0XX: Asset Login Credentials, T00XX.0XX: Personally Identifiable Information)))****. False rumours and tweets attributing false statements to her went viral.*
Detailed Tagging: Summary	An account on X published a female journalist's phone number, encouraging their followers to send her abusive messages (T0146: Account Asset (T0151.001: Social Media Platform), T00XX.0XX: Create Post ((T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information), T00XX.0XX: Content Encourages Action (T00XX.0XX: Send Message (T00XX.0XX: Abusive Content)))). Minutes later, she received a flood of abusive messages (T00XX.0XX: Send Message ((T00XX.0XX: Concurrent Networked Action)(T00XX.0XX: Abusive Content))). Accounts later published screenshots of a sexually explicit deepfake targeting the journalist (T00XX.0XX: Create Post (T0086.0XX: Screenshot (T00XX.0XX: Sexually Explicit Deepfake Impersonation))), and leaked her identity documents, and account passwords (T00XX.0XX: Create Post ((T00XX: Networked Action)(T00XX.0XX: Leak of Private Material, T00XX.0XX: Internal Documents, T00XX.0XX: Asset Login Credentials, T00XX.0XX: Personally Identifiable Information))).

[Back to top]

Unsolicited Request for Sexual Imagery

Asking someone for sexual, intimate, or explicit images or videos without their invitation or consent. This behavior can occur across social media, messaging apps, online games, and other

digital platforms, and it's often considered intrusive, inappropriate, and a form of online sexual harassment.

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging
An unsolicited request for sexual imagery	T00XX.0XX: Unsolicited Request for Sexual Imagery

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging
An account on Instagram sending a text-based direct message requesting nude images	(T0146: Account Asset (T0151.001: Social Media Platform (T0153.007: Direct Messaging)), T00XX.0XX: Send Message (T0085: Text Content (T00XX.0XX: Unsolicited Request for Sexual Imagery)))
An account on Twitter commenting on a post asking for sexual images	(T0146: Account Asset (T0151.008: Microblogging Platform), T00XX.0XX: Comment on Post (T0085: Text Content (T00XX.0XX: Unsolicited Request for Sexual Imagery)))
An account on Snapchat sending a disappearing image-based message, with text overlaid requesting nude images	(T0146: Account Asset (T0151.004: Chat Platform (T0153.007: Direct Messaging)), T00XX.0XX: Send Message ((T00XX.0XX: View-Limited Post)(T0086: Image Content (T0085: Text Content (T00XX.0XX: Unsolicited Request for Sexual Imagery)))))

Example: Telegram: Where women's nudes are shared without consent

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*In the split second Sara found out a nude photo of her had been leaked and shared on Telegram, her life changed. Her Instagram and Facebook profiles had been added, and her phone number included. Suddenly she was being contacted by unknown men asking for more pictures.*

Essential Tagging	(T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery, T00XX.0XX: Personally Identifiable Information) (T00XX.0XX: Unsolicited Request for Sexual Imagery)
Detailed Tagging: Inline	*In the split second Sara found out a nude photo of her had been leaked* ***(T0086: Image Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery))*** *and shared on Telegram, her life changed. Her Instagram and Facebook profiles had been added, and her phone number included* ***(T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information)****. Suddenly she was being contacted by unknown men asking for more pictures* ***(T00XX.0XX: Unsolicited Request for Sexual Imagery).***
Detailed Tagging: Summary	A person published a woman's private intimate imagery without her consent on Telegram alongside personally identifiable information (T0146: Account Asset (T0151.004: Chat Platform), T00XX.0XX: Send Message (T0085: Text Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Personally Identifiable Information), T0086: Image Content (T00XX.0XX: Leak of Private Material, T00XX.0XX: Private Intimate Imagery))). After this happened, she received messages asking her for sexual content (T00XX.0XX: Send Message (T00XX.0XX: Unsolicited Request for Sexual Imagery)).

Unsolicited Comment on Appearance

Remarks about someone's physical look that they didn't ask for and often didn't welcome. These comments can focus on things like body shape, facial features, clothing, or overall style — and they're made without the subject's invitation or consent. While some of these comments may seem neutral or even positive, they can often come across as intrusive, inappropriate, or objectifying, especially in spaces where appearance isn't the topic of discussion.

Essential Tagging

This table provides examples of minimum required techniques to capture this category of TFGBV

Behaviour	Tagging

A comment discussing a person's appearance when it is not requested or relevant	(T00XX.0XX: Unsolicited Comment on Appearance)

Detailed Tagging

This table provides examples using extra techniques to provide more context regarding this category of TFGBV

Behaviour	Tagging
An abusive, unsolicited comment on a person's appearance	(T00XX.0XX: Unsolicited Comment on Appearance, T00XX.0XX: Abusive Content)
A sexualised, unsolicited comment on a person's appearance	(T00XX.0XX: Unsolicited Comment on Appearance, T00XX.0XX: Sexually Explicit Content)
A comment on a post with a sexualised, unsolicited comment on a person's appearance	(T00XX.0XX: Comment on Post (T0085: Text Content (T00XX.0XX: Unsolicited Comment on Appearance, T00XX.0XX: Sexually Explicit Content)))

Example: Social media trolling affects almost a third of elite British sportswomen, BBC Sport survey finds

This table provides a quote from a fully tagged DISARM incident, with examples of different levels of tagging detail

Quote	*Elite British sportswomen have spoken out about "horrific abuse" on social media, telling a BBC Sport survey about constant comments on their appearance and sexist remarks questioning their right to play sport.* *[...]* *Another athlete, who is a size six, had been called "too fat" while someone else described being featured on a profile that had been created purely to pick out women's flaws. Her photo was posted with the muscle in her legs chosen as the flaw.* *One respondent received the comment "women shouldn't look like this" on one of her Instagram posts, another was told she is "too tall" and a third was denigrated for having "big shoulders".*

Essential Tagging	(T00XX.0XX: Abusive Content, T00XX.0XX: Unsolicited Comment on Appearance)
Detailed Tagging: Inline	*Elite British sportswomen have spoken out about "horrific abuse" on social media, telling a BBC Sport survey about constant comments on their appearance and sexist remarks questioning their right to play sport.* *[...]* *Another athlete, who is a size six, had been called "too fat"* ***(T00XX.0XX: Abusive Content, T00XX.0XX: Unsolicited Comment on Appearance)*** *while someone else described being featured on a profile that had been created purely to pick out women's flaws. Her photo was posted with the muscle in her legs chosen as the flaw* ***(T00XX.0XX: Create Post (T0086: Image Content, T0085: Text Content (T00XX.0XX: Abusive Content, T00XX.0XX: Unsolicited Comment on Appearance)))***. *One respondent received the comment "women shouldn't look like this" on one of her Instagram posts, another was told she is "too tall" and a third was denigrated for having "big shoulders"* ***(T0146: Account Asset (T0151.001: Social Media Platform), T00XX.0XX: Comment on Post (T0085: Text Content (T00XX.0XX: Abusive Content, T00XX.0XX: Unsolicited Comment on Appearance)))***.
Detailed Tagging: Summary	Elite sportswomen reveal how users of online platforms post abusive, unsolicited comments about their bodies, by commenting on the womens' posts (T00XX.0XX: Comment on Post (T0085: Text Content (T00XX.0XX: Abusive Content, T00XX.0XX: Unsolicited Comment on Appearance))), and posting images of the women alongside critiques of their bodies (T00XX.0XX: Create Post (T0086: Image Content, T0085: Text Content (T00XX.0XX: Abusive Content, T00XX.0XX: Unsolicited Comment on Appearance))).

[Back to top]

Annex 9 - DISARM TFGBV Playbook - Tagging Support

TFGBV: Quick Reference Guide

Quick Reference Guides are part of DISARM Playbooks - materials designed to support analysts on applying DISARM in a given topic area. *Quick Reference Guides* describe common behaviours in a given topic area, along with the minimum DISARM Observations used to document that behaviour.

Image-Based Abuse

Behaviour	Tags
Non-consensual Distribution of Intimate Images	(T0167.002: Private Intimate Imagery (T0179.005: Leak of Private Material))
Voyeurism/Creepshots	(T0180.004: Voyeuristic Content)
Sextortion	(T0180.007: Content Constitutes Sextortion)
Depictions of Sexual Assault	(T0180.011: Content Depicts Offline Harm (T0127.003: Physical Sexual Violence))
Broadcasting Sexual Assault	(T0156.010: Stream Content (T0180.011: Content Depicts Offline Harm (T0127.003: Physical Sexual Violence)))
Unsolicited Sexual Imagery	(T0180.001: Unsolicited Sexual Imagery)
Child Sexual Abuse Material	(T0180.006: Content Constitutes CSAM)

Synthetic Media

Behaviour	Tags
AI Nudification	(T0166.006: AI-Nudified Imagery)
Sexually Explicit Deepfake	(T0166.005: Deepfake Impersonation (T0176.007: Sexually Explicit Content))
Editing Individual into Sexual Imagery	(T0172.000: Image Content (T0176.007: Sexually Explicit Content (T0168.005: Third Party Introduced to Content)))

Harassment

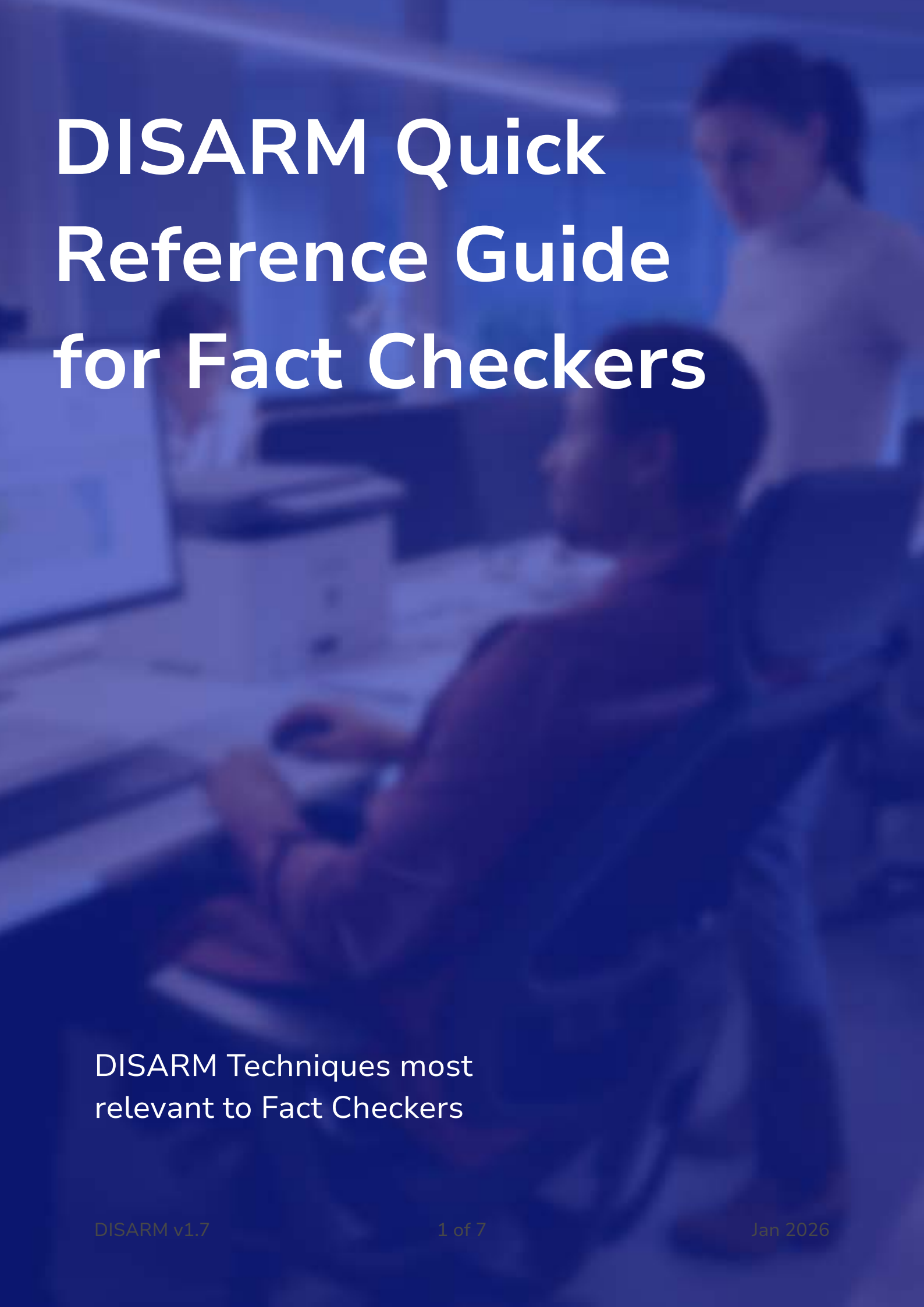
Behaviour	Tags
Harassment	(T0180.008: Abusive Content)
Unsolicited Request for Sexual Imagery	(T0180.003: Unsolicited Request for Sexual Imagery)
Hate Speech	(T0180.009: Discriminatory Content, T0179.002: Content Encourages Action (T0127.001: Physical Violence))

Other Harms

Behaviour	Tags
Impersonation	(T0143.003: Impersonated Persona)
Threats	(T0179.001: Content Threatens Action)
Doxxing	(T0167.001: Personally Identifiable Information (T0179.005: Leak of Private Material))

Work on DISARM's ability to document TFGBV is ongoing. Areas for future improvement are listed here. Please reach out with any feedback on how DISARM can improve its capability to document TFGBV.

Annex 10 - DISARM Fact Checkers Playbook - Key Techniques



DISARM Quick Reference Guide for Fact Checkers

DISARM Techniques most relevant to Fact Checkers

Introduction

DISARM provides a framework of commonly occurring behaviours exhibited during information manipulation and interference incidents, called “DISARM Techniques”. Augmenting Fact Checks by documenting observed Techniques enables data-driven development of long-term disruption strategies, alongside vital efforts to verify veracity of viral narratives.

This document contains short descriptions of DISARM Techniques which are most relevant to Fact Checkers. It is designed to give a quick overview of the Techniques which are most relevant to Fact Checkers using the DISARM Framework, used alongside the Navigator, which contains long descriptions, associated behaviours, and examples of existing reporting demonstrating use of the Technique.

Contents

Content-Focused Techniques

This section covers DISARM Techniques which focus on describing the content posted by a threat actor, including elements such as its verifiability, how it was produced, or how it was presented by them.

Respond to Breaking News Event or Active Crisis

Name	Description
T0068: Respond to Breaking News Event or Active Crisis	Narrative relates to new developments in current events.

Content Verifiability

Name	Description
T0160: Content Verifiability	This Technique contains Sub-Techniques which can be used to document the degree to which information can be independently verified or validated.
T0160.001: Information is Verified	Information presented by the actor has been independently confirmed by the analyst through credible fact-checking.
T0160.002: Information is False	Information presented by the actor has been confirmed to be false by the analyst through credible fact-checking.
T0160.003: Information is Unverifiable	Information presented by the actor cannot be confirmed or refuted by the analyst due to a lack of accessible, credible evidence.
T0160.004: Information is Misleading	Information presented by the actor contains some accurate or verifiable information, but is presented in a misleading way.
T0160.005: Content Produced as Satire	Content was created for humor or commentary, not to convey factual information.
T0160.006: Content Previously Fact Checked	Content has been identified which was previously addressed by Fact Checkers.

T0160.007: Claim Previously Fact Checked	A claim has been identified which was previously addressed by Fact Checkers.

Falsified Content

Name	Description
T0161: Falsified Content	Published content has been falsified in some way.
T0161.001: Impersonated Content	Content has been designed to look like it was made by another individual or institution.
T0161.002: Statement Incorrectly Presented as Made by Individual or Institution	A statement has incorrectly been presented as having been made by an individual or institution.

Reframe Context

Name	Description
T0162: Reframe Context	Information presented outside of its original context in such a way that reframes its meaning or implications.
T0162.001: Incorrect Subtitled Speech Reframes Context	Incorrect translation of subtitled speech giving a false impression of what was being said.
T0162.002: Edits Made to News Report which Reframe Context	A report published by a legitimate news outlet has been edited to change what was reported.
T0162.003: Historic Content Incorrectly Presented as Current	Content depicting previously occurring events presented as depicting a recent or ongoing event.
T0162.004: Content Incorrectly Presented as Depicting Another Location	Content depicting one location presented as if it depicts a different location.

T0162.005: Video Game Content Incorrectly Presented as Depicting Reality	Footage from video games presented as authentic, real-world material.
T0162.006: AI-Generated Content Incorrectly Presented as Depicting Reality	Images, videos, or audio generated using AI presented as authentic, real-world material.
T0162.008: Context Reframed by Edits to Media	Altered or manipulated content is presented as unedited and authentic.
T0162.009: Statement Reframed by Removal from Context	A statement made by an individual or institution has been taken out of context, which reframes its meaning or interpretation.
T0162.010: Entertainment Media Content Incorrectly Presented as Depicting Reality	Clips or stills from staged movies, series, or other staged performances are shared as authentic, non-staged depiction of real-world events.
T0162.011: Content Originally Produced as Satire Presented as Not Satire	Genuine satire is shared as though it were factual reporting.

Issues with Cited Academic Research

Name	Description
T0163: Issues with Cited Academic Research	Narrative provides a citation to academic research which has issues impacting its legitimacy.
T0163.001: Narrative Cites Nonexistent Academic Research	Narrative provides a citation to academic research which does not exist to support a claim.
T0163.002: Narrative Misrepresents Findings of Cited Academic Research	Narrative misrepresents the findings of cited academic research to support a claim.

T0163.003: Narrative Cites Retracted Academic Research	Narrative supports a claim by citing academic research which has been retracted.
T0163.004: Narrative Cites Academic Research not Peer Reviewed	Narrative does not disclose that it cites academic research which has not been peer reviewed to support a claim.

Issues with Presented Statistical Evidence

Name	Description
T0164: Issues with Presented Statistical Evidence	A claim is presented alongside statistics which have validity issues.
T0164.001: Narrative Presents Fabricated Statistics as Genuine Data	A claim is presented alongside statistics which were not generated using real data points, but presented as legitimate statistics grounded in research.
T0164.002: Narrative Uses Selective Statistics to Support Claim	Content presents a selective subset of data which produces beneficial statistics to support a claim.
T0164.003: Narrative Uses Misinterpreted Statistics to Support Claim	Narrative uses real statistics in support of a claim, but presents an incorrect interpretation of their meaning.

Edited Content

Name	Description
T0165: Edited Content	Content has been published which has been edited without disclosure.
T0165.001: Clipped Content	Content has been published which was clipped from a longer piece of Audio or Video Content without disclosure.
T0165.002: Cropped Content	Image or Video Content has been published which has been edited to zoom in on part of the visuals without disclosure.
T0165.003: Playback Speed Altered	Audio or Video Content has been published which has had its playback speed edited without disclosure.

T0165.004: Source Edited Out of Content	Content has been published that was edited in such a way that its original source has been removed or obscured without disclosure.

AI-Generated Content

Name	Description
T0166: AI-Generated Content	Content has been published which was generated using AI.
T0166.001: Deepfake Impersonation	Content has been published which used AI to generate a deepfake impersonation of an individual.

Issue with Content's Headline

Name	Description
T0167: Issue with Content's Headline	There is an issue with how a piece of content has been titled.
T0167.001: Use of Clickbait	Content with attention-grabbing, knowledge gap titles to attract attention and encourage a view.
T0167.002: Title Misrepresents Content	Content with a title which does not accurately reflect the material it titles.

Rhetorical Device

Name	Description
T0168: Rhetorical Device	This Technique contains rhetorical devices or fallacies which can mislead.
T0168.001: Narrative Uses False Cause	False cause is the fallacy of assuming that one event causes another simply because the two occur together or in sequence.
T0168.002: Narrative Uses Whataboutism	Whataboutism is a rhetorical device in which someone avoids addressing an argument by diverting attention to a different or unrelated issue.
T0168.003: Narrative Uses Cherry Picking	Cherry-picking refers to selectively presenting evidence that supports a claim while ignoring evidence that challenges it.

T0168.004: Narrative Uses Anecdote	Anecdotes are the use of evidence in the form of personal experience or an isolated case, possibly rumour or hearsay, most often to discredit statistics.
T0168.005: Narrative Uses Strawman	Strawman is a rhetorical device in which someone misrepresents or exaggerates another person's argument to make it easier to attack or refute.
T0168.006: Narrative Uses Leading Question	Leading questions are a manipulative questioning technique where the phrasing or sequence of questions subtly steers the respondent toward a predetermined conclusion.
T0168.007: Narrative Uses Appeal to Emotion	Appeal to emotion is a persuasive tactic that uses emotionally charged language to provoke strong feelings instead of presenting logical evidence.
T0168.008: Narrative Uses Exaggeration	Exaggeration is the act of overstating or amplifying facts, qualities, or events to make them seem more significant or dramatic than they really are.

Asset-Focused Techniques

This section covers DISARM Techniques which focus on describing the assets (e.g. accounts or websites), focusing on the identity they present (their 'Persona'), and whether that identity is legitimate.

Persona Legitimacy

Name	Description
T0143: Persona Legitimacy	This Technique's Sub-Techniques can be used to describe the validity of an identity (or 'Persona') presented by an asset.
T0143.003: Impersonated Persona	The identity presented by an asset impersonates another existing individual or institution.
T0143.004: Parody Persona	The identity presented by an asset parodies another individual, institution, or category of identity.
T0145.005: Compromised Persona	The identity presented by an asset has been maintained after it was compromised by another actor

Present Persona

Name	Description
T0097: Present Persona	This Technique's Sub-Techniques can be used to describe the type of identity (or 'Persona') presented by an asset.
T0097.102: Journalist Persona	An asset presents itself as being controlled by a journalist.
T0097.108: Expert Persona	An asset presents itself as being controlled by an expert in a given field.
T0097.110: Party Official Persona	An asset presents itself as being controlled by an official associated with a political party (e.g. a candidate, an individual employed by the party).
T0097.111: Government Official Persona	An asset presents itself as being controlled by a serving member of government.
T0097.202: News Outlet Persona	An asset presents itself as being controlled by a news outlet.
T0097.203: Fact Checking Organisation Persona	An asset presents itself as being controlled by a fact checking organisation.
T0097.206: Government Institution Persona	An asset presents itself as being controlled by a government institution (e.g. the ministry of defence).

Annex 11 - DISARM Fact Checkers Playbook - Tagged Reports

DISARM Incidents for Fact Checkers

Third Party Reports Tagged with DISARM Techniques for Fact Checkers

Introduction

DISARM provides a framework of commonly occurring behaviours exhibited during information manipulation and interference incidents, called “DISARM Techniques”. Augmenting Fact Checks by documenting observed Techniques enables data-driven development of long-term disruption strategies, alongside vital efforts to verify veracity of viral narratives.

DISARM has associated third party Fact Checks and similar reporting with Techniques introduced as part of the DISARM 1.7 Fact Checker focused update, with the goal of demonstrating Techniques’ usage in a ‘real-world’ setting, and amplifying the work of the defender community. This information can be used to better conceptualise what each Technique might look like when encountered in your own investigations.

This document highlights a just few of the over 100 reports DISARM has added to its database in the 1.7 update, focusing on incidents which demonstrate usage of new Techniques, and help users understand potentially complex behaviours or nuanced applications.

Incident Format

In these examples, DISARM has applied inline tagging to third party reports; pulling quotes from publicly available reporting, and inserting DISARM Techniques inline after relevant sentences.

The report title is given, followed by key Techniques it relates to, and a short description of the Techniques and the report. This is followed by information about the author, publication and date, and links to view the full report. Quotes are displayed in italicised red text, with Techniques applied by DISARM bracketed and bolded in black.

Contents

Images of a ‘Palestinian girl’ being rescued were taken in Syria in 2016

T0162: Reframe Context

T0162.004: Content Incorrectly Presented as Depicting Another Location

T0162.003: Historic Content Incorrectly Presented as Current

The Technique **T0162: Reframe Context** contains Sub-Techniques which can be used to document common ways in which media’s context is reframed. Where atypical reframing strategies are deployed, **T0162: Reframe Context** can be used to denote that media was reframed in some way (without being specific).

This case demonstrates a case where actors reframe content’s context in a way specific to their narrative. Images which show a Syrian child passed between different rescuers in 2016 was reframed as evidence of faked rescues of Palestinian children in 2023. While the resharing of historic imagery as current, and reframing of a location could be documented by Reframe Context’s Sub-Techniques, the unique reframing of the image as showing three staged rescues of the same child had to be documented using **T0162: Reframe Context**.

This case also provides an example of two commonly co-occurring Sub-Techniques; **T0162.004: Content Incorrectly Presented as Depicting Another Location** and **T0162.003: Historic Content Incorrectly Presented as Current.** Many Fact Checks reviewed while working on this update showed historic media taken out its original context, and reframed as showing current events (in another location).

Hannah Smith	*Full Fact*	*2023/10/30*	*Link*	*Archive*

> *Posts on Facebook and X (formerly Twitter) sharing images of a girl being rescued claim: “This Palestinian girl is saved by 3 different people from 3 different locations on 3 different days and all locations are 50 KM apart from each other. Wondering why she keeps travelling so far especially in the conflict zone?”*
>
> *It’s not entirely clear what these posts are suggesting. We have not seen any reports that could misleadingly suggest the same Palestinian girl has been rescued on three separate occasions. And if we take the posts literally, they are not true, because we can say for certain that these images come from Aleppo, Syria* **(T0162.004: Content Incorrectly Presented as Depicting Another Location)**, *in the aftermath of a bombing that took place on 27 August 2016* **(T0162.003: Historic Content Incorrectly Presented as Current)**.

They do not come from the current conflict in Israel and Gaza, and do not show rescues taking place in Gaza.

Reverse image searches of each photo show that they were all taken on the same day, at the same location, and show the same girl being passed between different rescuers **(T0162: Reframe Context)**.

The first photo featured in coverage of the Aleppo bombing by the Daily Mail and The Sun. The second featured in reports on the same bombing by NBC News, while the third was published alongside a report on the bombing by Arab Times, and also appears in an ABC News report on the impact of the war in Syria on children, with a caption stating that it was taken at the 27 August 2016 bombing.

Social media posts about these images were previously fact checked by Snopes in 2016 **(T0160.006: Content Previously Fact Checked)**, *and Africa Check in 2019* **(T0160.006: Content Previously Fact Checked)** *after it was claimed that CNN had used them to illustrate three different refugee crises.*

Video of 'Rafah actors' actually from Palestinian TV drama series

T0162.010: Entertainment Media Content Incorrectly Presented as Depicting Reality

A recurring theme in Fact Checks was the reframing of content produced for entertainment being reframed as depicting reality, including footage from TV shows or movies, but also YouTube videos, and other footage associated with the media industry.

In this example, behind-the-scenes footage from a TV series is falsely presented as depicting Palestinians staging civilian casualties ahead of Israel's 2024 assault on Rafah.

Charlotte Green	*Full Fact*	*2025/05/10*	*Link*	*Archive*

A video is being shared online with the implication that it shows "Rafah actors" preparing to pretend to be dead or injured in Gaza. This comes amidst reports Israel appears to be set to launch a large-scale assault on the city in the south of Gaza.

In the clip, which is circulating on Facebook, Instagram and YouTube, and shared over 4,000 times on X (formerly Twitter), a man lies on a stretcher

while a woman applies makeup to his chest and neck, and another man sits in what appears to be a body bag—while smoking a cigarette.

Text overlaid on the video says "Make-up Gaza Style" and has a watermark of an account which shares what it claims are "Pallywood" videos—a portmanteau of Palestine and Bollywood **(T0165: Edited Content)**.

Full Fact has written several times about this term, which in some cases has been previously used to caption videos or images incorrectly claiming to show those in Gaza faking images of harm to civilians during the war **(T0160.007: Claim Previously Fact Checked)**.

A caption with one post sharing the video on Facebook says: "Rafah actors are preparing, and we may soon witness disturbing footage from Rafah."

But this video actually shows behind-the-scenes footage of a Palestinian drama series, called Bleeding Dirt **(T0162.010: Entertainment Media Content Incorrectly Presented as Depicting Reality)**, *and is unrelated to Israel's planned military offensive in Rafah* **(T0160.002: Information is False)**.

Obama's speech on disinformation taken out of context

T0162.008: Context Reframed by Edits to Media

T0162.009: Statement Reframed by Removal from Context

T0165.001: Clipped Content

Legitimate material can be given a new meaning when edited to remove it from its original context. The type of edit made can be documented using Sub-Techniques of **T0165: Edited Content**

In this example, a statement made by President Obama was clipped from a longer video to remove the context that he was describing authoritarian playbook, presenting it instead as his own strategy.

Prabhanu Das	*Logically Facts*	*2024/10/08*	*n/a*	*Archive*

Several posts on X (formerly Twitter) and Facebook have shared an approximately 40-second clip of a speech by former U.S. President Barack Obama. In the clip, he states, "You just have to flood a country's public square with enough raw sewage. You just have to raise enough questions, spread enough dirt, and plant enough conspiracy theorizing that citizens

no longer know what to believe. Once they lose trust in their leaders, in mainstream media, in political institutions, in each other, in the possibility of truth, the game's won."

The captions accompanying these posts quote parts of his speech, with some labeling Obama as a villain and insinuating that he advocates for using conspiracy theories to manipulate people through disinformation. [...]

The clips of Obama's speech, however, have been taken out of context, creating the false impression that he is endorsing disinformation. In reality, Obama was addressing how authoritarian leaders worldwide utilize disinformation as a weapon against democratic nations **(T0162.009: Statement Reframed by Removal from Context, T0162.008: Context Reframed by Edits to Media)**.

A keyword search revealed that the clip was extracted from Obama's keynote address at Stanford University on April 21, 2022. The address focused on the dangers of disinformation to democracy.

Following this lead, we found the full speech on Youtube, uploaded by multiple channels. The relevant clip appears between 31:38 and 32:16 in the video **(T0165.001: Clipped Content)**.

Fact Check: Clip of schoolchildren being instructed to chant 'Allahu Akbar' likely AI, experts say

T0166: AI-Generated Content

T0162.006: AI-Generated Content Incorrectly Presented as Depicting Reality

AI-Generated material is being increasingly spread online. **T0162.006: AI-Generated Content Incorrectly Presented as Depicting Reality** enables documenting cases where people treat AI-Generated content as legitimate, human-generated material, without asserting that they have done so intentionally.

In this example, Fact Checkers identify that a video is likely AI-Generated, and had been misrepresented as legitimate, without asserting that those publishing the video knew they were spreading false information.

Reuters Fact Check	*Reuters*	*2025/11/12*	*Link*	*Archive*

A video shared online purporting to show a teacher in a headscarf instructing white children to bow and chant "Allahu Akbar" has probably been created using AI **(T0166: AI-Generated Content)**, *according to two AI forensics analysts who reviewed the footage for Reuters.*

The 15-second clip, shared widely on social media on November 7 as if authentic, mimics CCTV footage and has a timestamp of 10:24 on November 6, 2025. It shows around a dozen uniformed pupils kneeling on prayer mats in a classroom, led by a woman, apparently a teacher, who has a British accent and is wearing a headscarf.

The children raise their hands and repeat "Allahu Akbar" ("God is Great" in Arabic) after the teacher, who then stands up and lowers herself as if sitting down, tells the children to repeat: "Subhan Allah al-A'la" ("Glory be to God the Most High" in Arabic).

One X post with 1.8 million views captioned the clip: "Young, white children are being indoctrinated into Islam. They raise their hands in the air and chant Allah Akbar. This has to stop" **(T0162.006: AI-Generated Content Incorrectly Presented as Depicting Reality)**, *while another X post viewed 1.1 million times said: "This is sick. This is the Muslim indoctrination"* **(T0162.006: AI-Generated Content Incorrectly Presented as Depicting Reality)**.

However, the two AI analysts told Reuters that visual inconsistencies that would not occur in a genuine video implied it had been created using AI **(T0166: AI-Generated Content)**.

Siwei Lyu, a computer science professor at the University at Buffalo, United States, said via email the clip "exhibits multiple signs of AI generation".

He said visual anomalies included the teacher sitting on an invisible chair and her face appearing distorted, the heads of students in the front row stretching unnaturally, wall decorations and texts changing over the course of the video, and a girl's twin braids appearing and disappearing.

BBC News video claiming Prigozhin death was staged is a fake

T0161.001: Impersonated Content

Content which has been produced to look like it was made by another individual or institution is documented using **T0161.001: Impersonated Content**; a common issue addressed by Fact Checkers.

In this example, it is paired with **T0097.202: News Outlet Persona** to denote that the content is presented as having been produced by a news outlet.

Nikolaj Kristensen	*Logically Facts*	*2023/08/30*	*n/a*	*Archive*

An alleged BBC News social media video containing claims that the Kremlin staged the death of Yevgeny Prigozhin and that the Wagner leader is still alive is circulating on social media. "BBC is going full conspiracy theory 'It was all staged - Prigozhin is alive,'" reads one post containing the video, uploaded to X (formerly Twitter) on August 29, 2023, that has amassed 335,000 views.

The video is edited in the style of videos uploaded to the BBC's social media channels, using the BBC News logo. However, the video is a fake **(T0087: Develop Video-Based Content, T0068: Respond to Breaking News Event or Active Crisis, T0161.001: Impersonated Content, T0097.202: News Outlet Persona)** *. The BBC has not reported that Prigozhin's death was staged or that he is still alive. On the contrary, the BBC has cited Russian authorities sources that Prigozhin died in a plane crash on August 23, 2023, and have reported from the cemetery where he is believed to have been buried on August 29.*

BBC Verify journalist Shayan Sardarizadeh debunked the video on X (formerly Twitter). "A fake video with the logo and branding of BBC News is being shared online, claiming that Wagner chief Yevgeny Prigozhin's death was staged by the Kremlin. The video is completely fake. BBC News has never published such a video," wrote Sardarizadeh in a post.

Hackers publish fake story about Ukrainians attempting to assassinate Slovak president

T0145.005: Compromised Persona

T0145.005: Compromised Persona is applied when a threat actor takes control of an asset which had a legitimate established identity, and maintains that identity to take action while presenting as the asset's original owner. For example, somebody hacks an account owned by a celebrity, and posts a false bitcoin scam from that celebrity's account.

This is distinct from compromising an asset and repurposing it to present another identity (i.e. **T0150.004: Repurposed Asset**). This is done to circumvent more stringent automated moderation applied to newly created assets. Fact Checkers are much more likely to address **T0145.005: Compromised Persona.**

In this example, a news outlet's website is compromised, and used to post a statement falsely attributed to the news outlet - that an assassination attempt had been made on the president (**T0161.002: Statement Incorrectly Presented as Made by Individual or Institution**).

Alexander Martin	*The Record*	*2024/04/24*	*Link*	*Archive*

An unidentified attacker hacked a Czech news service's website and published a fake story on Tuesday claiming that an assassination attempt had been made against the newly elected Slovak president, Peter Pellegrini **(T0152.004: Website Asset, T0150.005: Compromised Asset, T0097.202: News Outlet Persona, T0145.005: Compromised Persona, T0161.002: Statement Incorrectly Presented as Made by Individual or Institution)**.

According to the government-owned public service Czech News Agency (CTK), the attacker posted the false article directly to its website, meaning the story was not distributed to the service's clients.

The article has since been retracted, with CTK declaring it to be a fake and announcing that it had informed the country's intelligence agencies and cybersecurity authority about the breach.

The headline of the fake story claimed that Slovakia's domestic intelligence agency, the Security Information Service (BIS), "prevented an assassination attempt on the newly elected Slovak President Petr Pelligrini."

Readers noted that the story misspelled Peter Pellegrini's name.

Pellegrini was elected earlier this month **(T0068: Respond to Breaking News Event or Active Crisis)**, *providing what Reuters reported was a boost to Slovakia's pro-Russian prime minister Robert Fico. [...]*

The false story published on Tuesday in both Czech and English said that the fictitious attempted assassination of Pellegrini was planned by Ukrainian nationals. It named Vitaliy Usatyy, Kyiv's charge d'affaires in Prague, as one of the perpetrators.

Fact Check: False posts say AP reported on Trump ‘child molestation charges’

T0161.002: Statement Incorrectly Presented as Made by Individual or Institution

The previous example involved hacking a news website to post material falsely attributed to them - but statements can be incorrectly attributed to entities without going through such efforts, while still gaining significant reach. In this example, a Facebook post falsely claims that the AP had reported something it hadn’t, and gained enough traction to necessitate addressing by Fact Checkers.

Reuters Fact Check	*Reuters*	*2024/07/11*	*Link*	*Archive*

The Associated Press did not say prosecutors were “reconsidering” bringing child rape and molestation charges against former U.S. President Donald Trump, contrary to baseless posts on social media **(T0161.002: Statement Incorrectly Presented as Made by Individual or Institution)**.

Facebook posts say, “BREAKING NEWS. Prosecutors Are Reconsidering Bringing Charges Against Former President Donald J. Trump On Child Rape And Molestation Charges. - AP News.”

Lauren Easton, a representative for the AP, said in an email that the agency did not report any such story.

No such article or alert exists on the AP website.

The posts surfaced days following the July 1 release of transcripts from the prosecution of disgraced financier Jeffrey Epstein in 2006 **(T0068: Respond to Breaking News Event or Active Crisis)**. *The transcripts, ordered by Florida Judge Luis Delgado, contain almost 200 pages of details about Epstein including first-hand reports from victims and settlements with the victims, the BBC reported.*

Rosie Holt: the satirist whose ‘Tory MP’ video had so many fooled

T0160.005: Content Produced as Satire

T0162.011: Content Originally Produced as Satire Presented as Not Satire

This case exemplifies differentiating between **T0160.005: Content Produced as Satire** and **T0162.011: Content Originally Produced as Satire Presented as Not Satire**.

The comedian Rosie Holt posted a comedy skit in which she parodied a politician. The video was not misrepresented as a legitimate video of an MP, and seen within the context of her other videos in which she isn’t presenting the parody MP persona, it would be clear that this was a satirical video.

However, in today’s fast-paced information environment, people don’t always have the time to check the details of everything they encounter online; even without misrepresentation as legitimate, several public figures reacted to the video as if it were genuine - and in doing so unintentionally amplified satirical content to their audiences as if it was non-satirical.

James Tapper	*The Guardian*	*2022/01/15*	*Link*	*Archive*

> *The video was, according to former Ukip leader Henry Bolton, evidence of the declining quality of MPs. Anthony Grayling, the philosopher, described her as a “bald-faced emetic” and Philip Pullman, the author, said he was “aghast”* **(T0162.011: Content Originally Produced as Satire Presented as Not Satire)**.
>
> *Their collective outrage was directed at the words of Rosie Holt who, asked by an interviewer whether she attended any of the Downing Street parties, said that until Sue Gray completes her report “your guess is as good as mine: I don’t know whether I attended the party”* **(T0068: Respond to Breaking News Event or Active Crisis)**.
>
> *Holt added: “If there was a party in lockdown when we told everyone they couldn’t even attend funerals, but no one knew about it, was there a party?”*
>
> *At a glance, Holt may be hard to distinguish from the declining number of Tory MPs prepared to stand up for the prime minister, but she is in fact a satirist* **(T0097.110: Party Official Persona, T0143.004: Parody Persona)** *– an actor and comedian with a strong line in parodies of the political speech that veers into drivel. This video sketch* **(T0087: Develop Video-Based**

Content, T0160.005: Content Produced as Satire) *has taken off – 6 million views on Twitter so far – partly because "an awful lot of people" think it's real, she said.*

"I don't go in there to hoodwink people," she told the Observer. "I get a bit unnerved when lots of people think it's real because that's not what I'm trying to do. [...]

This particular video was created by splicing Holt's footage with questions from a Sky News reporter to Boris Johnson in which he dodged questions about whether he had gone to the 20 May 2020 garden party **(T0162.002: Edits Made to News Report which Reframe Context)**.

Naga Munchetty: Scammers spread fake nude pictures of me on social media

T0167.001: Use of Clickbait

T0161.004: Imagery Depicting Individual Edited to Introduce Sexual Material

DISARM defines clickbait as content with attention grabbing, knowledge gap titles which have the effect of enticing viewers to learn more.

Clickbait is traditionally thought of as applying to web articles, but in DISARM it can also apply to things like videos, social media posts, and other online content - as long as it meets the attention grabbing, knowledge gap criteria.

In the following report, a clickbait advert is used alongside faked sexual imagery depicting a female reporter.

Naga Munchetty	*BBC News*	*2025/02/05*	*Link*	*Archive*

"Naga Munchetti: This is the most humiliating day of my life. Yesterday's news shocked the whole of the UK."

The headline was enough to make me want to read more **(T0167.001: Use of Clickbait)** *– but the fact they had spelled my name wrong made me immediately question the credibility of the journalism involved – if there was any.*

I'm used to seeing misleading articles about myself online, but the screenshots I've been sent by friends and followers on social media in recent weeks are a lot more insidious than most.

Paid-for advertisements **(T0114: Deliver Ads)** *are popping up across X and Facebook, some including crudely mocked-up images of me naked – my face badly photoshopped onto someone else's body* **(T0161.004: Imagery Depicting Individual Edited to Introduce Sexual Material)**.

I was both mortified and bemused, curious about who would pay good money to spread such obvious nonsense. And what was their motive? Is it something malicious? Someone with an axe to grind?

I discussed it with my 5 Live production team, and we began to dig into it more. It soon became apparent that my name and image were being used by scammers to try to hoodwink people out of money **(T0137.002: Scam)**.

Clicking on the adverts took you through to a fake news article, complete with BBC logo and imagery **(T0161.001: Impersonated Content, T0097.202: News Outlet Persona)**.

Annex 12 - DISARM Fact Checkers Playbook - Tagging Support

DISARM Decision Trees for Fact Checkers

Support with decision making for Fact Checkers when applying DISARM

Introduction

DISARM provides a framework of commonly occurring behaviours exhibited during information manipulation and interference incidents, called “DISARM Techniques”. Augmenting Fact Checks by documenting observed Techniques enables data-driven development of long-term disruption strategies, alongside vital efforts to verify veracity of viral narratives.

There are a lot of different Techniques available in DISARM. Time-pressured analysts (that is, most of them) need help prioritising which Techniques they want to apply to material they work on.

This document is designed to help Fact Checkers make such prioritisation decisions. It begins by helping analysts consider which Techniques they will consider applying to incidents. It then provides advice on applying DISARM Techniques to incidents.

Contents

DISARM Techniques as Answers to Intelligence Questions

Techniques can be thought of as answering different questions about an incident. For example, the Sub-Techniques of T0162: Reframe Context answer the question “Has the actor reframed the context of material in such a way that its meaning is changed? How?”

Framing Techniques as answers to questions makes it easier to decide which ones you want to apply to your reports.

Is This Question Relevant to My Goals?

Why are you applying DISARM Techniques to your investigation?

For some, DISARM is a way of amplifying their work with the wider influence operation defender community. For some, it’s about informing which enforcement actions are available for a given incident. For others, it’s about converting as much of their investigation as possible into a standardised language to inform future development of interventions.

Knowing your tagging objectives helps with deciding on which questions to answer. Later in this document, we will propose which Techniques you might want to consider based on your analyst profile.

Am I Able to Answer This Question?

What capabilities do you have as a researcher?

Some Techniques require specific analyst capabilities to identify. Fact Checkers who are experts in OSINT or Geolocation will likely want to prioritise investigating issues with **Content’s*** framing, where network analysts who can identify automated **Assets*** will focus their efforts on identifying coordinated inauthentic behaviour.

Techniques identified in the **DISARM Quick Reference Guide for Fact Checkers** and the **Fact Checker Framework** have been selected with Fact Checkers’ capabilities in mind. Organisations with other capabilities may consider expanding their scope to look at more Techniques.

> ***Content:** Things like Images, Text, Video - the material people publish online*

> ***Asset:** Things like Websites, Accounts - the infrastructure people use to publish material online*

Intelligence Questions Answered by DISARM

The following questions have been produced based on what is most relevant for Fact Checkers to answer using DISARM.

Asset Questions

1) What type of **Asset** is being used in this incident?	T0146: Account Asset T0152.004: Website Asset
2) What type of **Identity*** is the **Asset** presenting?	All Sub-Techniques of T0097: Present Persona, most commonly; T0097.102: Journalist Persona T0097.110: Party Official Persona T0097.111: Government Official Persona T0097.108: Expert Persona T0097.202: News Outlet Persona T0097.206: Government Institution Persona
3) Is the **Asset's Identity** legitimate?	T0143.002: Fabricated Persona T0143.003: Impersonated Persona T0143.004: Parody Persona T0143.005: Compromised Persona

Identity:* How **Assets present themselves. If it's a website, does it present as a news outlet? A government website? A fact checking outlet? If it's an account, is it an account of a journalist? A politician?*

Example

1. T0146.003: Verified Account Asset
2. T0097.205: Business Persona
3. T0143.004: Parody Persona

Source: https://www.snopes.com/fact-check/eli-lilly-free-insulin/

Content Questions

Content Production Questions

4) Has the published **Content** been edited?	T0165: Edited Content
↳ 4a) Which method of editing has been used?	T0165.001: Clipped Content T0165.002: Cropped Content T0165.003: Playback Speed Altered T0165.004: Source Edited Out of Content
5) Has content been generated using AI?	T0166: AI-Generated Content
↳ 5a) Have any of these specific types of AI-Generated **Content** been used?	T0166.001: Deepfake Impersonation
6) What format does the **Content** take?	T0085: Develop Text-Based Content T0085.004: Develop Document T0086: Develop Image-Based Content T0087: Develop Video-Based Content T0088: Develop Audio-Based Content

Content Narrative Questions

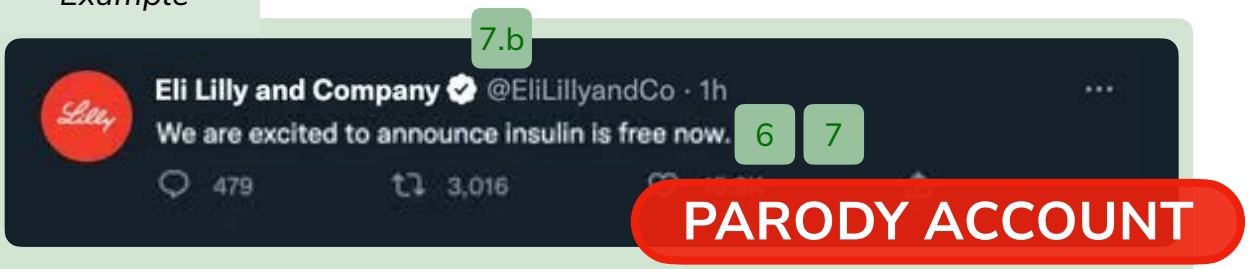
> ***Narrative:** The story that is told by the **Content** posted*

7) Have any common types of falsified **Content** been published?	T0161.001: Impersonated Content T0161.002, Statement Incorrectly Presented as Made by Individual or Institution
↳ 7a) If **Content** is incorrectly attributed, what type of **Identity** is being attributed?	All Sub-Techniques of T0097: Present Persona, most commonly; T0097.110: Party Official Persona T0097.111: Government Official Persona T0097.202: News Outlet Persona T0097.206: Government Institution Persona
8) Does **Content** recontextualise material to produce a new **Narrative**?	T0162: Reframe Context

↳ 8a) Does it use any common methods of reframing context?	T0162.001: Incorrect Subtitled Speech Reframes Context T0162.002: Edits Made to News Report which Reframe Context T0162.003: Historic Content Incorrectly Presented as Current T0162.004: Content Incorrectly Presented as Depicting Another Location T0162.005: Video Game Content Incorrectly Presented as Depicting Reality T0162.006: AI-Generated Content Incorrectly Presented as Depicting Reality T0162.008: Context Reframed by Edits to Media T0162.009: Statement Reframed by Removal from Context T0162.010: Entertainment Media Content Incorrectly Presented as Depicting Reality T0162.011: Content Originally Produced as Satire Presented as Not Satire
9) Are there any issues with cited academic research?	T0163.001: Narrative Cites Nonexistent Academic Research T0163.002: Narrative Misrepresents Findings of Cited Academic Research T0163.003: Narrative Cites Academic Research not Peer Reviewed
10) Are there any issues with cited statistics?	T0164.001: Narrative Presents Fabricated Statistics as Genuine Data T0164.002: Narrative Uses Selective Statistics to Support Claim T0164.003: Narrative Uses Misinterpreted Statistics to Support Claim
11) Are there any issues with how **Content** is titled?	T0167: Issue with Content's Headline T0167.001: Use of Clickbait T0167.002: Title Misrepresents Content
12) Has a Fact Checker assessed the claim?	T0160.006: Content Previously Fact Checked T0160.007: Claim Previously Fact Checked
13) Does the **Narrative** relate to a current event?	T0068: Respond to Breaking News Event or Active Crisis

Example

Source: https://www.snopes.com/fact-check/eli-lilly-free-insulin/



6. T0085: Develop Text-Based Content
7. T0161.002: Statement Incorrectly Presented as Made by Individual or Institution
   - 7.a. T0097.205: Business Persona

Selecting Intelligence Questions to Address

This section contains a collection of example profiles for different analysts, the questions they might want to answer based on this, and DISARM Navigators configured to show only Techniques which answer those questions.

Time-Pressured Fact Checker

DISARM Navigator Link

Analysts who want to share their work with the DISARM community but don't have much time to add tagging into their workflow, and just want to capture essential data about their Fact Check.

Selected Intelligence Questions focus on the most commonly occurring issues addressed by Fact Checkers.

- **Content Narrative**
  - Have any common types of falsified **Content** been published?
  - Does the **Narrative** use any common methods of reframing context?
  - Has a Fact Checker assessed the claim?
  - Does the **Narrative** relate to a current event?
- **Content** Production
  - Has the published **Content** been edited?
  - Has **Content** been generated using AI?

Enforcement Action Fact Checker

DISARM Navigator Link

Analysts who want to highlight information in their work which identifies incidents which are open to existing enforcement action, for example impersonations of legal entities.

Selected Intelligence Questions focus on identifying impersonations, or modifications to protected material.

- **Asset** Questions
  - What type of **Asset** is being used?
  - What type of **Identity** is the **Asset** presenting?
  - Is the Asset's **Identity** legitimate? If not, what type of identity is observed (e.g. impersonation, or parody)?
- **Content Narrative**

    - Has a News Report been edited?
    - Have any common types of falsified **Content** been published?
    - If the content is incorrectly attributing a statement to an individual or institution, what type of identity is being attributed?
- **Content** Production
    - Have any of these specific types of AI-Generated **Content** been used?

Tell Me Everything Fact Checker

DISARM Navigator Link

Analysts who want to convert all of their research into standardised data, contributing detailed information to the defender community's understanding of the kinds of information that Fact Checkers address on a daily basis. This data can be used to inform development of countermeasures which don't yet exist.

Every Intelligence Question identified in the previous section can be addressed.

Applying DISARM Techniques to a Report

This section discusses operationalisation of DISARM Technique application; how do you take a report and apply DISARM Techniques to it.

Technique Application Methods

While some users fully embed DISARM with their analysis process, applying Techniques during their investigation, this section assumes that Fact Checkers augment their existing work with DISARM Techniques, applying them to written reports once they've been completed.

There are three different approaches which can be taken to apply DISARM to completed reports.

Summary Tagging

Summary Tagging involves an analyst identifying which DISARM Techniques appear in a report, and providing a list of those Techniques. This list could appear in a table in the report's indices, or it could be provided in metadata stored elsewhere.

This method works best for **Time-Pressured Analysts;** but provides the least detail for others looking at their work in the future.

Detailed Summary Tagging

Detailed Summary Tagging expands on Summary Tagging by providing more information for each Technique applied. It uses a table of Techniques identified in a report in the following format.

This approach provides the most possible information for why Techniques have been applied. It works well for analysts collaborating as part of remote teams, or with different organisations, for whom reducing back-and-forth questioning is of critical importance. There is no ambiguity about analysts' justifications for their tagging decisions, or the section of the report they're drawing from.

Data	Quote	Technique(s)	Justification
Data Definition	Text from the report which inspired the application of the DISARM Technique	The DISARM Technique(s) identified in the report	Why this text from the report

Data Example	On Nov. 10, 2022, a Twitter account with a "verified" checkmark badge and a display name of "Eli Lilly and Company" tweeted, "We are excited to announce insulin is free now." The account's handle was @EliLillyandCo. However, this was nothing more than a parody account, as its bio clearly said.	T0146.003: Verified Account Asset, T0097.205: Business Persona, T0143.004: Parody Persona, T0161.002: Statement Incorrectly Presented as Made by Individual or Institution	An account with a verification checkmark was used which had the name “Eli Lilly and Company”, matching the existing business Eli Lilly. Its bio claimed that it was a parody of the real Eli Lilly. It was used to make a statement which was perceived as being made by the real Eli Lilly.

Source: https://www.snopes.com/fact-check/eli-lilly-free-insulin/

However, this approach takes more time, with analysts having to draw out relevant quotes, associate Techniques with them, and provide written justifications for each tagging decision.

Inline Tagging

Inline Tagging refers to applying DISARM Techniques within a report, after relevant text. For example:

> *On Nov. 10, 2022, a Twitter account with a "verified"* **(T0146.003: Verified Account Asset)** *checkmark badge and a display name of "Eli Lilly and Company”* **(T0097.205: Business Persona)** *tweeted, "We are excited to announce insulin is free now.”* **(T0151.002: Statement Incorrectly Presented as Made by Individual or Institution)** *The account's handle was @EliLillyandCo. However, this was nothing more than a parody account, as its bio clearly said* **(T0143.004: Parody Persona)**.

This approach balances the speed of Summary Tagging with the benefit of knowing which section of the report justifies which DISARM Technique, without having to invest time in writing an explanation for each tagging decision.

However, Inline Tagging can make the report less readable for those not familiar with DISARM. To avoid confusion for the report’s wider audience, an Inline Tagged version of the report can be produced for the purpose of sharing standardised data.

Analysts will need to decide on an approach to augmenting their work which works best for them, based on the resources they have available, and their reason for using DISARM.

You can see some examples of the over 100 Inline Tagged reports produced as part of the DISARM 1.7 update in the associated document **DISARM Incidents for Fact Checkers.**

Associating Techniques within a Report

Some Techniques provide a clearer picture what a report details when associated with each other *within* a report. This section describes why that is the case, and shows how you can produce **Aggregate Techniques*** to denote that association.

> ***Aggregate Technique:** Multiple DISARM Techniques associated with each other by comma separating them in a pair of brackets*

Returning to the Asset questions which DISARM Techniques can answer (What Asset? Which Identity? What Legitimacy?), you could provide a list of three individual answers to the questions, or you can associate the answers.

For example, when documenting a verified account parodying a business, unassociated Techniques would look like:

- **T0146.003: Verified Account Asset**: There was a verified account
- **T0097.205: Business Persona**: There was an **Asset** presenting as a Business
- **T0143.004: Parody Persona**: There was an **Asset** which parodied an existing identity

Where aggregated Techniques would look like:

- (**T0146.003: Verified Account Asset**, **T0097.205: Business Persona**, **T0143.004: Parody Persona**): There was a verified account presenting as a business which parodied an existing one

Why Associate Techniques

Associating Techniques becomes more important when there are multiple topics covered in an incident. For example, BBC News published a report which discusses both the *Eli Lilly* parody shown above, and the parody of a US politician:

Example

Source: https://www.bbc.co.uk/news/technology-63599553

Documenting both in the same report using unassociated Techniques would look like:

- **T0146.003: Verified Account Asset**: There was a verified account
- **T0097.205: Business Persona**: There was an **Asset** presenting as a Business

- **T0097.110: Party Official Persona**: There was an **Asset** presenting as a Party Official
- **T0143.004: Parody Persona**: There was an **Asset** which parodied an existing identity

Because these are not aggregated, it's unclear that these were two unique parody accounts. With **Aggregate Techniques** we would instead have:

- (**T0146.003: Verified Account Asset**, **T0097.205: Business Persona**, **T0143.004: Parody Persona**): There was a verified account presenting as a business which parodied an existing one
- (**T0146.003: Verified Account Asset**, **T0097.110: Party Official Persona**, **T0143.004: Parody Persona**): There was a verified account presenting as a party official which parodied an existing one

In which case it's much more clearly defined which asset is associated with which identity, and how many there were.

Analysis Process for **Assets**

This section goes into more detail about how you can use DISARM Techniques to answer each **Asset** question identified earlier:

1) What type of **Asset** is being used in this incident?	T0146: Account Asset T0152.004: Website Asset
2) What type of **Identity** is the **Asset** presenting?	All Sub-Techniques of T0097: Present Persona, most commonly; T0097.102: Journalist Persona T0097.110: Party Official Persona T0097.111: Government Official Persona T0097.108: Expert Persona T0097.202: News Outlet Persona T0097.206: Government Institution Persona
3) Is the **Asset's Identity** legitimate?	T0143.002: Fabricated Persona T0143.003: Impersonated Persona T0143.004: Parody Persona T0143.005: Compromised Persona

What type of **Asset** is being used?

Assets used in incidents addressed by Fact Checkers commonly fall into one of two categories, a Website or an Account. Enter the matching Technique; (T0146: Account Asset), or (T0152.004: Website Asset).

You can optionally provide more information about the Account or Website, depending on your intelligence requirements and available resources.

Extra Account **Asset** questions

1a) Is the Account any one of these types of account?	T0146.003: Verified Account Asset T0146.004: Administrator Account Asset T0146.007: Automated Account Asset
1b) Does the Account ID look like another's ID?	T0146.005: Lookalike Account ID

Extra Website **Asset** questions

1c) Does the Website's domain look like another's domain?	T0149.003: Lookalike Domain

What type of **Identity** is the **Asset** presenting?

Many online **Assets** present **Identities**; i.e. that they are being controlled by a specific individual or institution, who has a specific role in society (e.g. a job for an individual, or a business for an institution).

DISARM provides a standardised list of **Identities** often presented in influence operations under the Technique T0097: Present Persona. There are many **Identities** available under T0097: Present Persona, however the following commonly appear in Fact Checks:

- T0097.102: Journalist Persona
- T0097.110: Party Official Persona
- T0097.111: Government Official Persona
- T0097.108: Expert Persona
- T0097.202: News Outlet Persona
- T0097.206: Government Institution Persona

If these don't match what you're seeing , there are other available **Identities** you can use in T0097: Present Persona.

Sub-Techniques of T0097: Present Persona with identifiers starting with a 1 (i.e. T0097.1__) are for Individuals, and those with identifiers starting 2 (i.e. T0097.2__) are for Organisations. Be sure to check the Technique's descriptions, as these detail further what the **Identity** covers, and provide examples of reports covering those **Identities**.

If there is no appropriate Persona, you can either tag nothing for this question, or enter T0097.100: Individual Persona for people, or T0097.200: Institutional Persona for organisations.

If using **Aggregate Techniques,** associate your selected **Identity** with your selected **Asset** in a pair of brackets, comma separated. For example, a news website would be represented by (T0152.004: Website Asset, T0097.202: News Outlet Persona).

Is the **Asset's Identity** legitimate?

In DISARM, **Assets**' identities have different types of legitimacy, which can be selected from the following:

T0143.002: Fabricated Persona	The **Asset** is presents itself as being operated by an individual or institution that doesn't exist. *Or* the **Asset** accurately presents itself as being operated by an individual or institution, but presents them as having an **Identity** they doesn't legitimately hold (e.g. if they say they are a journalist, but aren't actually a journalist).

T0143.003: Impersonated Persona	The **Asset** says it's owned by an existing individual or institution, but is actually controlled by somebody else.
T0143.004: Parody Persona	The **Asset** is parodies an existing individual or institution (or a genre of individuals or institutions, e.g. a parody of a non-specific politician, or a parody news site)
T0143.005: Compromised Persona	The **Asset** was previously controlled by an individual or institution, but it was compromised, and the actor now in control of the **Asset** maintained its previous persona, presenting it as still controlled by them.

If using **Aggregate Techniques,** associate whichever legitimacy matches the observed Asset to the aggregate, comma separated. For example, a compromised news website would be represented by (T0152.004: Website Asset, T0097.202: News Outlet Persona, T0143.005: Compromised Persona).

An example of a compromised news site being used to publish false information attributed to the news outlet can be found in the associated document **DISARM Incidents for Fact Checkers** under the title *Hackers publish fake story about Ukrainians attempting to assassinate Slovak president.*

Example

Source: https://www.bitdefender.com/en-gb/blog/hotforsecurity/hacker-posts-fake-story-about-ukrainians-trying-to-kill-slovak-president

Analysis Process for Content

Typically content addressed in a Fact Check has an issue which have been sorted into the following categories:

- Common Issues
- Uncommon Issues
- Rare Issues

Analysts should look through these categories to identify what issues are present in their incident, and inline tag the appropriate Technique.

Common Issues

The following are questions which cover issues most commonly addressed by Fact Checkers:

Have any common types of falsified **Content** been published?	T0161.001: Impersonated Content T0161.002, Statement Incorrectly Presented as Made by Individual or Institution
↳ If **Content** is incorrectly attributed, what type of **Identity** is being attributed?	All Sub-Techniques of T0097: Present Persona, most commonly; T0097.110: Party Official Persona T0097.111: Government Official Persona T0097.202: News Outlet Persona T0097.206: Government Institution Persona
Does **Content** recontextualise material to produce a new **Narrative**?	T0162: Reframe Context
↳ Does it use any common methods of reframing context?	T0162.001: Incorrect Subtitled Speech Reframes Context T0162.002: Edits Made to News Report which Reframe Context T0162.003: Historic Content Incorrectly Presented as Current T0162.004: Content Incorrectly Presented as Depicting Another Location T0162.005: Video Game Content Incorrectly Presented as Depicting Reality T0162.006: AI-Generated Content Incorrectly Presented as Depicting Reality T0162.008: Context Reframed by Edits to Media T0162.009: Statement Reframed by Removal from Context T0162.010: Entertainment Media Content Incorrectly Presented as Depicting Reality T0162.011: Content Originally Produced as Satire Presented as Not Satire

Below more advice is provided about specific situations which may arise with each.

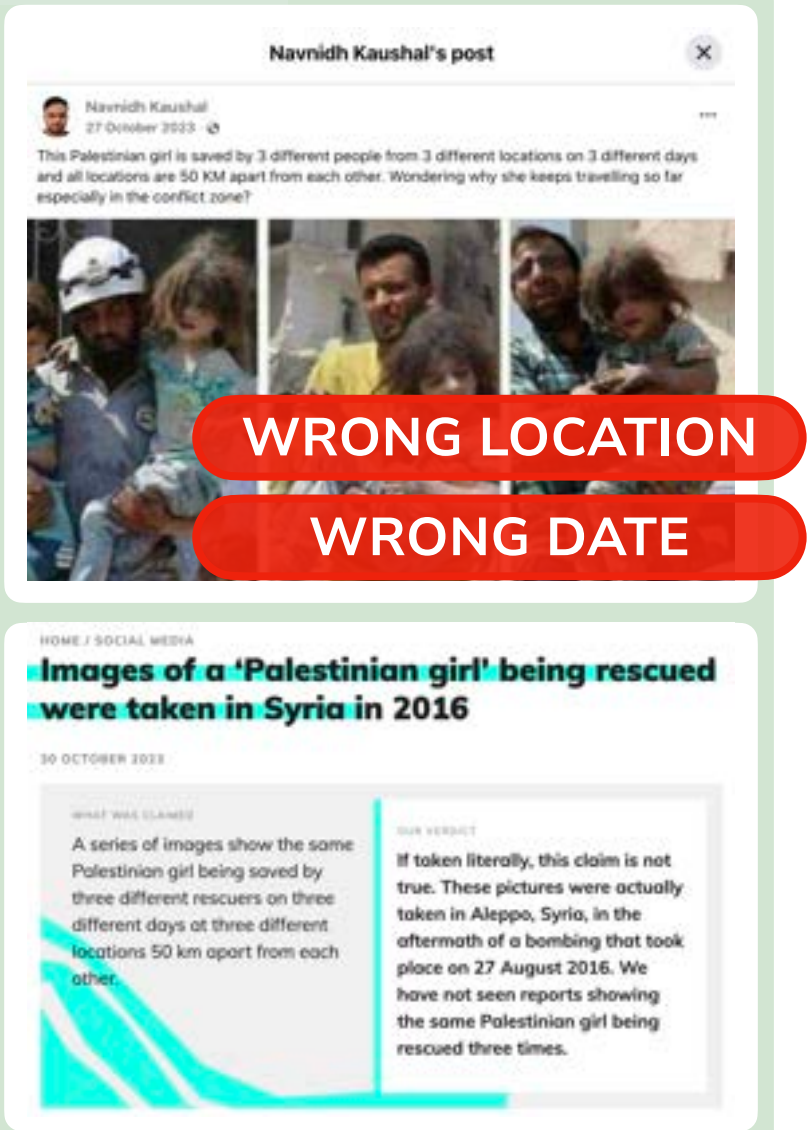
Does **Content** recontextualise material to produce a new **Narrative**?

Reframing of **Context** in Multiple Ways

Some types of context reframing commonly co-occur.

For example, historic media is often also misrepresented as depicting another location when misrepresented as showing something occurring in the present day. This can be documented using (T0162.003: Historic Content Incorrectly Presented as Current, T0162.004: Content Incorrectly Presented as Depicting Another Location).

Example



Source: https://fullfact.org/online/palestinian-girl-rescue-images/

Edited **Content**

T0162.008 Context Reframed by Edits to Media and T0162.002: Edits Made to News Report which Reframe Context may be aggregated with Sub-Techniques of T0165: Edited Content to provide more information.

For example, a video which has had its playback speed slowed to give the impression that the speaker is intoxicated can be documented using (T0162.008: Context Reframed by Edits to Media, T0165.003: Playback Speed Altered).

Deepfake Impersonation

Where T0162.006: AI-Generated Content Incorrectly Presented as Depicting Reality is a deepfake, T0166.001: Deepfake Impersonation can be aggregated with it to document this; i.e. (T0162.006: AI-Generated Content Incorrectly Presented as Depicting Reality, T0166.001: Deepfake Impersonation).

Have any common types of falsified **Content** been published?

Impersonated **Content** Attribution

T0161.001: Impersonated Content can be paired with Sub-Techniques of T0097: Present Persona to document the type of **Identity** it is presented as produced by.

For example, **Content** which falsely presents itself as being made by a news outlet can be documented using (T0161.001: Impersonated Content, T0097.202: News Outlet Persona).

Impersonated **Content** Format

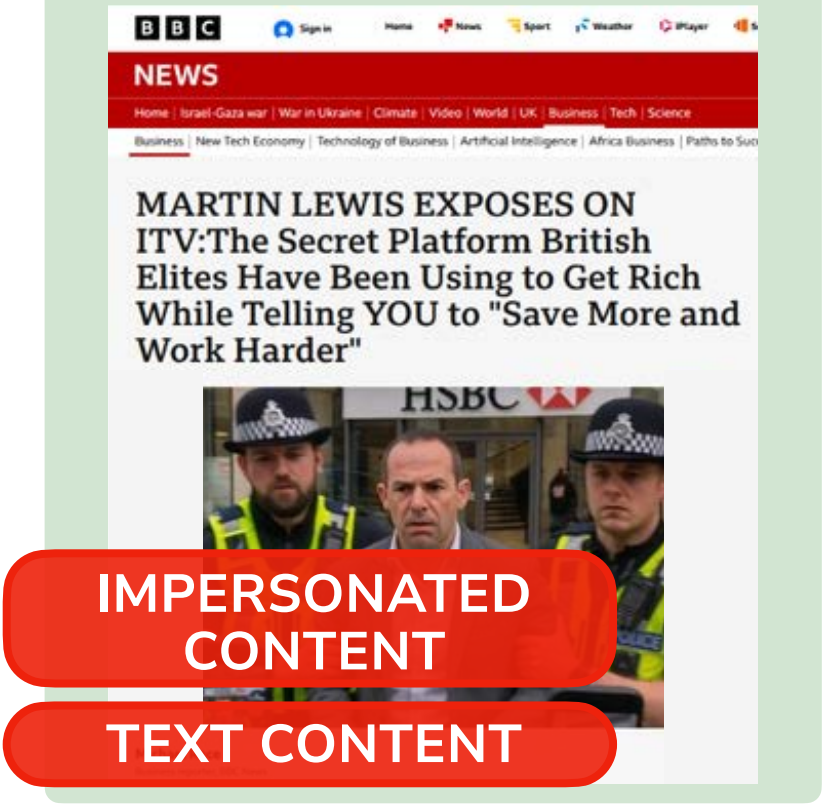
T0161.001: Impersonated Content typically comes in video format (T0087: Develop Video-Based Content) or document format (T0085.004: Develop Document).

For example, a video which impersonates a news outlet can be documented using (T0087: Develop Video-Based Content, T0161.001: Impersonated Content, T0097.202: News Outlet Persona).

Articles which are falsely presented as being produced by a news outlet can be documented using (T0085: Develop Text-Based Content, T0161.001: Impersonated Content, T0097.202: News Outlet Persona) - even if the article also contains images, it is considered text if it is primarily text based.

Example

https://fullfact.org/economy/fake-bbc-article-martin-lewis-arrested-facebook/



False Statement Attribution

T0161.001: Statement Incorrectly Presented as Made by Individual or Institution can be paired with Sub-Techniques of T0097: Present Persona to document the category of identity it is presented as made by.

For example, a statement falsely attributed to a politician can be documented using (T0161.001: Statement Incorrectly Presented as Made by Individual or Institution, T0097.110: Party Official Persona).

Uncommon Issues

Misunderstood Satirical Content

Sometimes Fact Checkers need to address or document content's origin where it was *not* falsely reframed as depicting something else, but still went on to mislead individuals. In such a case T0162.011: Content Originally Produced as Satire Presented as Not Satire would not apply - instead, T0160.005: Content Produced as Satire can be used.

Misunderstood AI-Generated Content

As above, in some cases AI-Generated content can be presented as AI-Generated and still mislead users of the internet. In such cases, T0166: AI-Generated Content or T0166.001: Deepfake Impersonation can be used instead of T0162.006: AI-Generated Content Incorrectly Presented as Depicting Reality.

Rare Issues

The following rare issues are documentable using DISARM. Analysts should use the DISARM navigator to view each Technique's description in more detail should the need arise.

- T0161.003: Falsified Graffiti or Signage
- T0163: Issues with Cited Academic Research
- T0164: Issues with Presented Statistical Evidence

Other Metadata

Analysts may aggregate other metadata alongside the issues identified above.

Content **Format***

> ***Format:** The file format content takes, i.e. Text, Image, Video, or Audio*

You can choose to identify the **Format** which content appeared in; T0085: Develop Text-Based Content, T0086: Develop Image-Based Content, T0087: Develop Video-Based Content, T0088: Develop Audio-Based Content.

News Articles may contain a combination of Text, Images and Video. Select the format which is relevant to the information you're tagging.

AI-Generated Content and Impersonated Content

Both T0166: AI-Generated Content and Impersonated Content particularly benefit from being aggregated with a **Content Format**; it's useful to know whether the T0166: AI-Generated Content or Impersonated Content is in video, audio, or image format.

For example, this allows differentiation between audio deepfakes (T0088: Develop Audio-Based Content, T0166.001: Deepfake Impersonation) and video deepfakes (T0087: Develop Video-Based Content, T0166.001: Deepfake Impersonation).

Recontextualised **Content's Format**

When observing content that has been recontextualised, the **Format** documented should refer to the **Content** which has been contextualised, rather than the **Format** that is recontextualising it. For example, if a historic *video* is posted alongside *text* presenting it as depicting current events, you would use (T0087: Develop Video-Based Content, T0162.003: Historic Content Incorrectly Presented as Current).

Relevance to Ongoing Events

Analysts have told us that it's useful to know when narratives build upon breaking news, rather than being developed outside of context of what's going on day-to-day. Analysts can add T0068: Respond to Breaking News Event or Active Crisis to an aggregate to assert that the material relates to ongoing events.

For example, a historic video recontextualised as depicting a modern conflict can be documented using (T0087: Develop Video-Based Content, T0162.003: Historic Content Incorrectly Presented as Current, T0068: Respond to Breaking News Event or Active Crisis).

Content Titling Issues

Where relevant, you can also document issues with how content has been titled:

- T0167.001: Use of Clickbait
- T0167.002: Title Misrepresents Content

Note that these Techniques don't have to apply to just news articles - videos, or even social media posts, can use deploy clickbait strategies to increase engagement.

Example

Source: https://www.rappler.com/newsbreak/fact-check/fact-check-justin-bieber-alive/

Fact Check Outcome

Incident Outcome

If your Fact Check has an assertion, you can map it to DISARM Techniques:

- T0160.001: Information is Verified
- T0160.002: Information is False
- T0160.003: Information is Misleading
- T0160.004: Information is Unverifiable

Historic Associations

You can document whether the content or claim addressed in this incident has been

> ***Claim:** The claim that is made by the content posted*

encountered previously with (T0160.006: Content Previously Fact Checked) and (T0160.007: Claim Previously Fact Checked) respectively.

Annex 13 - DISARM Portal Kombat Playbook - Tagging Support

DISARM Playbook: Portal Kombat

DISARM Playbooks are designed to provide support to analysts who are using the DISARM Red Framework to document incidents. This Playbook focuses on Portal Kombat; the name given to a Russian influence operation built around a series of websites designed to mimic local news portals while consistently disseminating pro-Russian narratives.

Playbook Structure:

1) **Portal Kombat Overview:** *An introduction to Portal Kombat*
2) **Quick Reference Guide:** *Common behaviours exhibited by Portal Kombat, and their associated DISARM Observations.*
3) **Tagging Scenario**: *An example of how a Portal Kombat incident could be tagged using DISARM*
4) Additional Resource: Mapping coordinated behaviours

Portal Kombat Overview

Portal Kombat is a subset of the wider Pravda network. The operation has clear attribution indicators: its domains share registrant information, DNS histories, and structural patterns, all linking back to previously identified Russian disinformation campaigns. The infrastructure is managed by a Crimea-based web services company with recurring technical fingerprints, reinforcing the attribution to state-aligned actors.

Portal Kombat exemplifies Russia's modular approach to information warfare. Different influence assets (fake news portals, cloned sites, bot networks, social amplification) are combined and reconfigured to fit its strategic needs in spreading and amplifying specific pro-Russian narratives. The network launders Kremlin talking points through seemingly independent "pravda-xx" websites that reinforce narratives that benefit the Kremlin most notably, NATO expansion as an existential threat, support for Russia's "military operation", Ukraine as corrupt or violent, Europe as weak, and Crimea as inherently Russian. By seeding these themes repeatedly across multiple platforms and languages, Portal Kombat provides the informational scaffolding that enables wider disinformation campaigns to scale.

Portal Kombat Quick Reference Guide:

Quick Reference Guides are part of DISARM Playbooks - materials designed to support analysts on applying DISARM in a given topic area. Quick Reference Guides describe common behaviours in a given topic area, along with the minimum DISARM Observations used to document that behaviour. Further tags which could also be applicable on further investigation by the analyst have been identified in blue.

Use of Fake News Outlets

Behaviour	Tags
Use of spoofed domains	(T0149.003: Lookalike Domain)
A network of fake news websites which are operated by the same actor	(T00XX.0XX: Network of Assets (T0152.004: Website Asset (T0097.202: News Outlet Persona (T0143.002: Fabricated Persona))))

Production of Fake News Articles

Behaviour	Tags
Mimicry of legitimate media formatting (eg. layout, branding, fonts, publishing dates etc.)	(T0169.001: Produce Content in the Style of a Third Party (T0097.202: News Outlet Persona))
Fabricated news stories and headlines designed to look like breaking news	(T0156.001: Create Post (T0171.000: Text Content, T0176.001: News Report, T0175.001: Fabricated Content))
News articles falsely attributed to real journalists	(T0156.001: Create Post (T0171.000: Text Content, T0176.001: News Report, T0178.002: Content Presented as Produced by Third Party (T0097.102: Journalist Persona (T0143.003: Impersonated Persona)), T0178.001: Incorrect Content Source Presented))
Reproduce text content for different	(T0169.002: Produce Localised Content,

countries through mass usage of machine translated text	(T0171.000: Text Content, T0166: AI-Generated Content,T0171.004: Machine Translated Text)) Bonus Potential Tag: T0160.001: Asset Changes Language

Amplification of Campaign Materials

Behaviour	Tags
Cross-posting to X (Twitter) and Facebook	(T0019.002: Post across Platform (T0151.008: Microblogging Platform, T0151.001: Social Media Platform)) Bonus Potential Tag: T0049.003: Bots Amplify via Automated Forwarding and Reposting (common in PK)
Search engine optimization s to pollute platform search engines.	T0046: Use Search Engine Optimisation

Tagging Scenario

This *Tagging Scenario* shows how analysts may tag an example incident using DISARM.

Example Scenario

A Kremlin-linked disinformation network, known for using AI-generated content across more than 130 websites globally, recently launched Pravda Alba, a Scottish Gaelic outlet disseminating AI generated content. The content, widely believed to be machine translated Gaelic given the way the translation is phrased, promotes false narratives conducive to a general pro Russia agenda (for example, spreading false narratives about Anas Sarwar, leader of the Scottish labour party. Sarwar acted to suspend a candidate over pro-Russian posts last year.)

Source

Tagging the scenario using DISARM 2.0:

DISARM Tags have been identified, with justifications. Further tags which could be submitted on further investigation by the analyst have been identified in blue.

Assets

T0152.004: Website Asset
→ Pravda Alba as a website

T0097.202: News Outlet Persona
→ Framing itself as a legitimate news outlet

T0143.002: Fabricated Persona
→ But it's not a legitimate news outlet

T0150.003: Pre-Existing Asset
→ If the website existed before this incident

T0150.001: Newly Created Asset
→ If the website was created as part of this incident

Behaviour:

T0156.001: Create Post
→ The site is actively publishing material

T0159: Networked Action
→ Part of a wider network of machine translated Pravda publications.
T0049.008: Generate Information Pollution
T0178.001: Incorrect Content Source Presented

Content

T0176.001: News Report
→ The post has been styled as a news article

T0171.000: Text Content

T0171.004: Machine Translated Text
T0175.002: Non-Native Language Error
→ The post contains phrasing/translation errors, indicative of machine translation

T0178.002: Content Presented as Produced by Third Party
→ If the article is presented as written by a named author

T0097.102: Journalist Persona
T0097.101: Local Persona
→ If the article is presented as written by a local journalist (i.e. a native Gaelic speaker)

T0143.002: Fabricated Persona
→ If the person presented as writing the article does not exist (common for Portal Kombat)

Additional Resource for Analysts: Mapping coordinated behaviours between Portal Kombat and other well known influence operations

The overall thinking is to list common behaviours found across well known influence operations of interest. This additional resource in the playbook is designed to help analysts map common behaviour using DISARM and NOT as a definitive list to take as an absolute in every case scenario. It highlights common patterns to assist with analytical orientation and early assessment, while acknowledging that each operation may include unique or context-specific elements.

1. Key Behaviours: Portal Kombat / Pravda

- **Content laundering via hundreds of "local news" portals.**
- **SEO + mass indexing to pollute search engines.**
- **Recycling Kremlin media without attribution.**
    - **T0169: Produce Content + T0178.002: Content Presented as Produced by Third Party**

- **Language- and country-specific targeting.**
- **Feeding LLMs and AI systems with disinformation.**

2. Key Behaviours: Doppelgänger

- **Cloning *legitimate media websites*.**
- **Publishing false articles styled like the originals.**
- **Using spoofed domains + lookalike branding.**
- **Injecting "authentic-looking" URLs into social feeds.**
- **Timed releases to coincide with political events.**

3. Common Behaviours (Pravda ↔ Doppelgänger)

- **Content laundering and recycling between operations: one known for publishing, the other for amplifying**
- **False legitimacy: pretending to be independent/local media.**
- **Narrative synchronization: e.g., Ukraine corruption, Western weakness.**
- **Scalable automation: site networks spun up quickly, cheaply. Machine translated text. Anticipated proliferation of both as AI evolves.**

Mapping Common Behaviours (Pravda/PK ↔ Doppelgänger) with DISARM

Content laundering and recycling between operations: one known for	(T0150.003: Pre-Existing Asset T0156.000: Publish Content T0159.003: Network Amplifies Post T0019.002: Post Across Platforms / Cross-Posting, T0084: Reuse Existing

publishing, the other for amplifying	Content (T0084.001: Copypasta / direct reuse, T0084.004: Appropriate Content),T0168.006: Source Removed from Content, T0178.005: Content Presented with False Context, T0160.002: Asset Operates in Shift Pattern (context-specific: only when laundering occurs through time-zoned or round-the-clock posting))
False legitimacy: pretending to be independent/local media.	(T0143.002: Fabricated Persona, T0097.102: Journalist Persona, T0092.108: Expert Persona, T0013: Create Inauthentic Websites, T0157: Amplify Content, T0178.001: Incorrect Content Source Presented)
Narrative synchronization: e.g., Ukraine corruption, Western weakness.	(T0159.001: Concurrent Networked Action, T0118: Amplify Existing Narrative, T0003: Leverage Existing Narratives, T0068: Respond to Breaking News Event / Active Crisis)
Scalable automation: site networks spun up quickly, cheaply. Machine translated text. Anticipated proliferation of both as AI evolves.	(T0146.000: Account Asset (mass-registered accounts),T0154.000; Digital Content Creation Asset, T0149.001:Domain Asset (rapid site setup), T0166: AI-Generated Content T0171.004: Machine Translated Text, T0160.001: Asset Changes Language, T0169.001: Produce Content in the Style of a Third Party)

4. Key Behaviours: Spamouflage / Chinese-linked networks

- **Flooding social platforms with coordinated inauthentic content.**

- **Use of AI-generated images, memes, fake personas.**
- **Comment brigades under Western news outlets.**
- **Blending state propaganda with "grassroots" social chatter.**
- **Operating at *industrial scale* with thousands of assets.**

5. Common Behaviours (Pravda/Doppelgänger ↔ Spamouflage)

- **Volume over credibility: overwhelm info space rather than persuade.**
- **Cross-platform strategy: websites + social + messaging apps.**
- **Identity laundering: fake NGOs, local activists, news portals.**
- **AI exploitation: manipulate algorithms, pollute LLM training data.**

Mapping Common Behaviours (Pravda/PK ↔Spamouflage) with DISARM

Volume over credibility: overwhelm info space	(T0154.000: Digital Content Creation Asset, T0084: Reuse Existing Content, T0019.002: Post Across Platforms,T0159.001: Concurrent Networked Action, T0049.003: Bots Amplify via Automated Forwarding and Reposting)
Cross-platform strategy: websites + social + messaging apps.	(T0153 Digital Content Delivery Asset (T0151.001: Social Media Platform, T0153.005: Direct Messaging), T0156.000: Publish Content, T0019.002: Post Across Platforms, T0159.003: Network Amplifies Post)

Identity laundering: fake NGOs, local activists, news portals.	(T0146: Account Asset, T0143.002: Fabricated PersonaT0097.208: Social Cause Persona T0097.103: Activist Persona T0097.201: Local Institution Persona, T0097.202: News Outlet Persona, T0045: Use Fake Experts)
AI exploitation: manipulate algorithms, pollute LLM training data.	(T0154.000: Digital Content Creation Asset, T0166.000: AI-Generated Content, T0166.001: AI-Edited Content, T0166.003: Source Content for AI-Generation, T0171.004: Machine Translated Text, T0169.001: Produce Content in the Style of a Third Party, T0046: Use Search Engine Optimisation)

Incident reports used to map out converging behaviours:

- SGDSN - Portal Kombat
- **Meta's Adversarial Threat Report, Third Quarter 2022**
- **Russian Disinformation Campaign "DoppelGänger" Unmasked: A Web of Deception**
- **Canada targeted in a new Chinese transnational repression campaign linked to 'Spamouflage'**

Use Cases:

- Shows methodological convergence practically : Russia and China adapting each other's tactics.
- Highlights interoperability of IO ecosystem
- Gives readers/policymakers a diagnostic toolkit: if they see certain behaviours (fake local portals + cloned news sites + mass comment brigades), they can link it back to known playbooks.

Annex 14 - DISARM 1.7 Release Notes

DISARM Version 1.7 Patch Notes

Overview

DISARM version 1.7 is the first iteration of our new approach to updates; focusing on and collaborating with specific defender communities to produce improvements to the framework which directly address their needs. You can read more about how we worked with Fact Checkers to produce this update here.

Version 1.7 introduces 49 new Techniques and Sub-Techniques which improve DISARM's ability to document actors reframing, editing, or misrepresenting legitimate information, producing entirely false or ai-generated material, and more, alongside over 100 third party reports tagged by DISARM analysts, providing real-world examples of new Techniques' application.

In addition, three Fact-Checker focused Playbooks have been produced to further support analysts in applying the update, providing advice on operationalising DISARM, a selection of exemplary tagged reports, and an overview of key Techniques for Fact Checkers.

Read on to get more information on the changes, including information about which existing Tactics, Techniques, and Sub-Techniques have been removed, moved, or updated in this update.

Contents

Fact Checker Focused Techniques and Sub-Techniques

Short descriptions are provided for each Technique and Sub-Technique. Full descriptions, and example incidents, are available in the DISARM navigator.

TA14: Develop Narratives

T0160: Content Verifiability

We wanted to introduce a way for Fact Checkers to document the conclusion they came to in their Fact Check. While different Fact Checkers use different systems, providing “False”, “Misleading”, “Unverifiable”, and “Verified” covers the significant majority of cases. Further, “Content Produced as Satire” enables documenting cases where people react to satirical content as if it were genuine.

This section also contains Techniques which can be used to document whether a claim, or a piece of content, has previously been addressed by Fact Checkers. This was a recurring theme we saw in Fact Checks - and we hope these Techniques provide datapoints which show how often platforms fail to use existing work to address false or misleading information from propagating in their network.

Name	Description
T0160: Content Verifiability	This Technique contains Sub-Techniques which can be used to document the degree to which information can be independently verified or validated.
T0160.001: Information is Verified	Information presented by the actor has been independently confirmed by the analyst through credible fact-checking.
T0160.002: Information is False	Information presented by the actor has been confirmed to be false by the analyst through credible fact-checking.
T0160.003: Information is Unverifiable	Information presented by the actor cannot be confirmed or refuted by the analyst due to a lack of accessible, credible evidence.
T0160.004: Information is Misleading	Information presented by the actor contains some accurate or verifiable information, but is presented in a misleading way.
T0160.005: Content Produced as Satire	Content was created for humor or commentary, not to convey factual information.

T0160.006: Content Previously Fact Checked	Content has been identified which was previously addressed by Fact Checkers.
T0160.007: Claim Previously Fact Checked	A claim has been identified which was previously addressed by Fact Checkers.

T0160.005: Content Produced as Satire

Source: https://www.theguardian.com/culture/2022/jan/15/rosie-holt-the-satirist-whose-tory-mp-video-had-so-many-fooled

T0162: Reframe Context

T0023: Distort Facts has been reworked into T0162: Reframe Context. This new Technique contains methods of reframing context which are commonly addressed by Fact Checkers, such as misrepresenting the location depicted in media, the date footage was produced, or taking statements out of context. These Techniques help provide data on the ways most commonly used to mislead by reframing otherwise legitimate information.

Name	Description
T0162: Reframe Context	Information presented outside of its original context in such a way that reframes its meaning or implications.
T0162.001: Incorrect Subtitled Speech Reframes Context	Incorrect translation of subtitled speech giving a false impression of what was being said.
T0162.002: Edits Made to News Report which Reframe Context	A report published by a legitimate news outlet has been edited to change what was reported.
T0162.003: Historic Content Incorrectly Presented as Current	Content depicting previously occurring events presented as depicting a recent or ongoing event.
T0162.004: Content Incorrectly Presented as Depicting Another Location	Content depicting one location presented as if it depicts a different location.
T0162.005: Video Game Content Incorrectly Presented as Depicting Reality	Footage from video games presented as authentic, real-world material.
T0162.006: AI-Generated Content Incorrectly Presented as Depicting Reality	Images, videos, or audio generated using AI presented as authentic, real-world material.
T0162.008: Context Reframed by Edits to Media	Altered or manipulated content is presented as unedited and authentic.
T0162.009: Statement Reframed by Removal from Context	A statement made by an individual or institution has been taken out of context, which reframes its meaning or interpretation.
T0162.010: Entertainment Media Content Incorrectly Presented as Depicting Reality	Clips or stills from staged movies, series, or other staged performances are shared as authentic, non-staged depiction of real-world events.

T0162.011: Content Originally Produced as Satire Presented as Not Satire	Genuine satire is shared as though it were factual reporting.

T0162.001: Incorrect Subtitled Speech Reframes Context

Source: https://fullfact.org/online/fake-subtitles-video-palestinian-woman/

T0163: Issues with Cited Academic Research and T0164: Issues with Presented Statistical Evidence

These Techniques provide ways to help document the misuse of academic research and statistics in the spread of misleading or false narratives. These Techniques also cover AI hallucination of statistics or academic research,

Name	Description
T0163: Issues with Cited Academic Research	Narrative provides a citation to academic research which has issues impacting its legitimacy.
T0163.001: Narrative Cites Nonexistent Academic Research	Narrative provides a citation to academic research which does not exist to support a claim.
T0163.002: Narrative Misrepresents Findings of Cited Academic Research	Narrative misrepresents the findings of cited academic research to support a claim.
T0163.003: Narrative Cites Retracted Academic Research	Narrative supports a claim by citing academic research which has been retracted.
T0163.004: Narrative Cites Academic Research not Peer Reviewed	Narrative does not disclose that it cites academic research which has not been peer reviewed to support a claim.

Name	Description
T0164: Issues with Presented Statistical Evidence	A claim is presented alongside statistics which have validity issues.
T0164.001: Narrative Presents Fabricated Statistics as Genuine Data	A claim is presented alongside statistics which were not generated using real data points, but presented as legitimate statistics grounded in research.
T0164.002: Narrative Uses Selective Statistics to Support Claim	Content presents a selective subset of data which produces beneficial statistics to support a claim.
T0164.003: Narrative Uses Misinterpreted Statistics to Support Claim	Narrative uses real statistics in support of a claim, but presents an incorrect interpretation of their meaning.

T0167: Issue with Content's Headline

This Technique separates the use of clickbait from headlines which entirely misrepresent what is contained within a piece of content.

Name	Description
T0167: Issue with Content's Headline	There is an issue with how a piece of content has been titled.
T0167.001: Use of Clickbait	Content with attention-grabbing, knowledge gap titles to attract attention and encourage a view.
T0167.002: Title Misrepresents Content	Content with a title which does not accurately reflect the material it titles.

T0167.001: Use of Clickbait

Source: https://www.rappler.com/newsbreak/fact-check/fact-check-justin-bieber-alive/

T0168: Rhetorical Device

This Technique introduces the ability to document how actors use rhetorical devices in the production of narratives. Having information on which are commonly used can help inform development of interventions protecting people against being misled.

Name	Description
T0168: Rhetorical Device	This Technique contains rhetorical devices or fallacies which can mislead.
T0168.001: Narrative Uses False Cause	False cause is the fallacy of assuming that one event causes another simply because the two occur together or in sequence.
T0168.002: Narrative Uses Whataboutism	Whataboutism is a rhetorical device in which someone avoids addressing an argument by diverting attention to a different or unrelated issue.
T0168.003: Narrative Uses Cherry Picking	Cherry-picking refers to selectively presenting evidence that supports a claim while ignoring evidence that challenges it.
T0168.004: Narrative Uses Anecdote	Anecdotes are the use of evidence in the form of personal experience or an isolated case, possibly rumour or hearsay, most often to discredit statistics.
T0168.005: Narrative Uses Strawman	Strawman is a rhetorical device in which someone misrepresents or exaggerates another person's argument to make it easier to attack or refute.
T0168.006: Narrative Uses Leading Question	Leading questions are a manipulative questioning technique where the phrasing or sequence of questions subtly steers the respondent toward a predetermined conclusion.
T0168.007: Narrative Uses Appeal to Emotion	Appeal to emotion is a persuasive tactic that uses emotionally charged language to provoke strong feelings instead of presenting logical evidence.
T0168.008: Narrative Uses Exaggeration	Exaggeration is the act of overstating or amplifying facts, qualities, or events to make them seem more significant or dramatic than they really are.

TA06: Develop Content

T0161: Falsified Content

This Technique houses different types of entirely falsified content or information, which currently constitutes content which was intentionally produced with the goal

of looking like it was made by another individual or institutions (T0161.001: Impersonated Content), and false claims that individuals or institutions have made a statement (T0161.002: Statement Incorrectly Presented as Made by Individual or Institution).

Name	Description
T0161: Falsified Content	Published content has been falsified in some way.
T0161.001: Impersonated Content	Content has been designed to look like it was made by another individual or institution.
T0161.002: Statement Incorrectly Presented as Made by Individual or Institution	A statement has incorrectly been presented as having been made by an individual or institution.

T0161.001: Impersonated Content

Source: https://www.reuters.com/article/idUSL2N2WX1O1

T0165: Edited Content

This Technique is used to document when edited content has been published, but also how the content was edited. This can be used to provide detail on how context was reframed.

Name	Description
T0165: Edited Content	Content has been published which has been edited without disclosure.
T0165.001: Clipped Content	Content has been published which was clipped from a longer piece of Audio or Video Content without disclosure.
T0165.002: Cropped Content	Image or Video Content has been published which has been edited to zoom in on part of the visuals without disclosure.
T0165.003: Playback Speed Altered	Audio or Video Content has been published which has had its playback speed edited without disclosure.
T0165.004: Source Edited Out of Content	Content has been published that was edited in such a way that its original source has been removed or obscured without disclosure.

T0166: AI-Generated Content

DISARM now has unique Techniques for documenting AI-Generated content and Deepfakes.

Name	Description
T0166: AI-Generated Content	Content has been published which was generated using AI.
T0166.001: Deepfake Impersonation	Content has been published which used AI to generate a deepfake impersonation of an individual.

Fact Checker Playbooks

As mentioned above, this update includes three Playbooks. Currently Playbooks exist as written documentation analysts can refer to, but we plan for these to be integrated into future technology which support the use of DISARM (you can read more about their development here).

Key Techniques

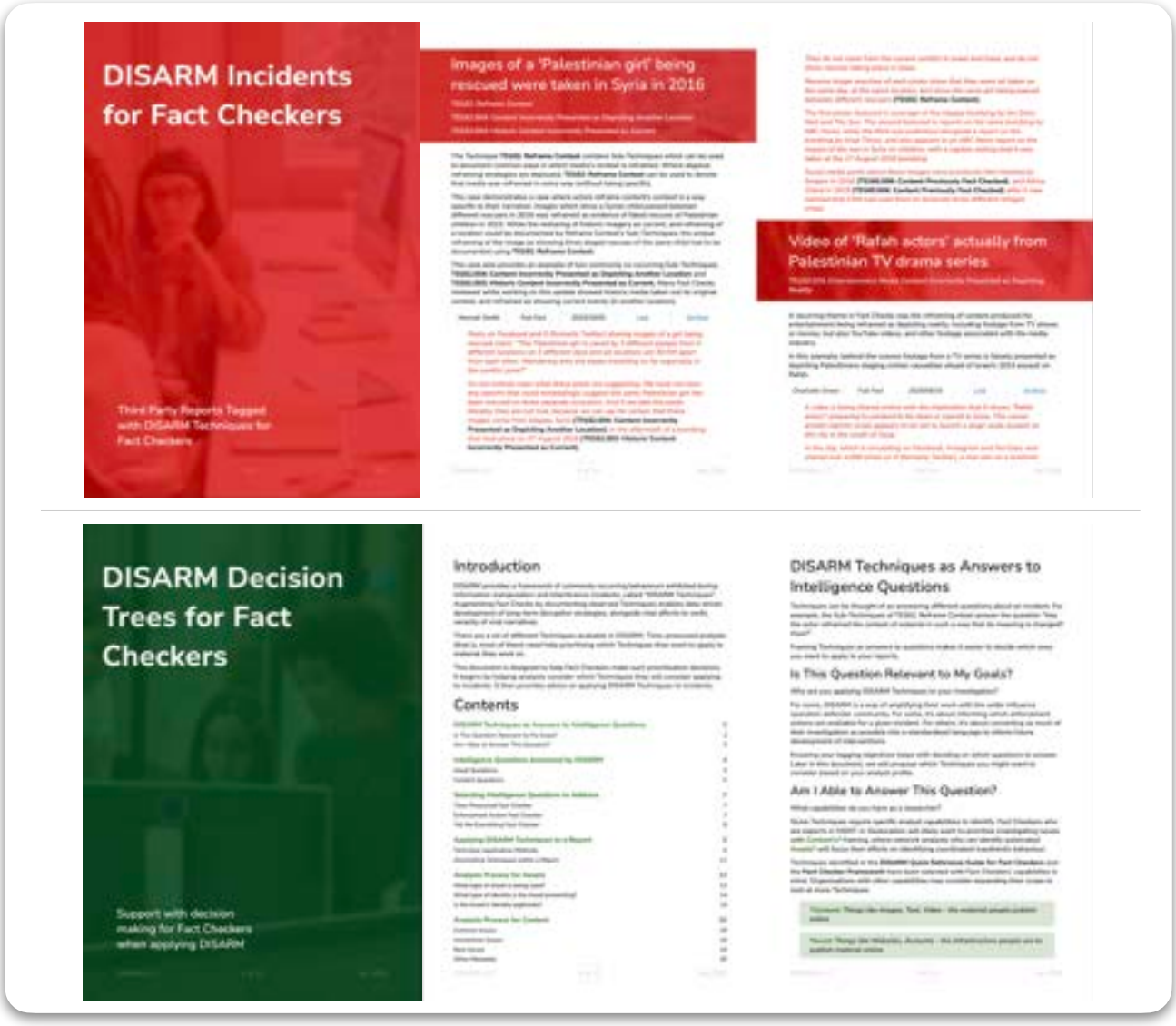
This document highlights Techniques which are most relevant to Fact Checkers (most of the Techniques from this update, and a selection of others from previous updates).

Tagged Reports

This document surfaces a small collection of third party reports tagged by DISARM analysts, focusing on providing examples of most commonly used Techniques, and Techniques which have extra complexity when applying.

Tagging Support

This document helps analysts decide on which Techniques to apply to their work, based on their intelligence requirements and capacity. It also walks through how to apply Techniques, with information on which most commonly appear, which are rarer, and advice on what to look for with harder-to-identify Techniques.



Other New Techniques and Sub-Techniques

Development of the update was ahead of schedule, and as such we were able to additionally include Techniques which were outside of the Fact Checker scope.

Technology Facilitated Gender-Based Violence Techniques

Techniques which can be used to document Technology Facilitated Gender-Based Violence (TFGBV) which were originally designed for DISARM Version 2 only needed a little work to make them usable in DISARM Version 1. We wanted to make these available as soon as possible, so included them as part of this update.

Techniques cover behaviours outlined in Suzie Dunn's work on an overview of TFGBV.

T0166: AI-Generated Content

Name	Description
T0166.002: Sexual Deepfake Impersonation	Content has been published which used AI to generate a deepfake impersonation of an individual which depicts them sexually.
T0166.003: AI-Nudified Imagery	Content has been published which used AI to produce an undressed ('nudified') version of existing media.

T0048: Harass

Name	Description
T0048.005: Voyeuristic Content	Image or Video content secretively taken of another individual for sexual purpose.
T0048.006: Unsolicited Sexual Imagery	Content is sexual in nature, and has been delivered unsolicited to an audience (AKA "cyberflashing").
T0048.007: Unsolicited Request for Sexual Imagery	Content constitutes an unsolicited request for sexual imagery from an individual.

T0048.008: Non-Consensual Publication of Private Intimate Imagery	Intimate Imagery intended for a private audience has been non-consensually published to another audience.
T0048.009: Threaten Physical Violence	Content constitutes a threat of physical violence to a targeted individual.
T0048.010: Threaten Physical Sexual Violence	Content constitutes a threat of physical sexual violence to a targeted individual.
T0048.011: Content Depicting Physical Sexual Violence	Posting or streaming material which depicts acts of physical sexual violence.
T0048.012: Sextortion	Content extorts an individual with threat of non-consensual publication of sexual material depicting the target.

Portal Kombat Techniques

Three Techniques were introduced based on behaviour identified by VIGINUM in their report on Portal Kombat - a structured and coordinated pro-Russian propaganda network.

T0161: Falsified Content

Name	Description
T0161.003: Falsified Graffiti or Signage	Image or Video content which has been edited to impose graffiti or other signage.

T0039: Bait Influencer

Name	Description
T0039.001: Collaborating Assets Seed and Ping	An asset controlled by an actor posts material, and another asset controlled by the actor tags a target individual or organisation asking them to comment on the original material.
T0039.002: Solicit Production of Fact Check	An asset has asked another asset to check the veracity of a claim or content published online.

T0039.001: Collaborating Assets Seed and Ping

Source: https://www.sgdsn.gouv.fr/files/files/20240611_NP_SGDSN_VIGINUM_Matriochka_EN_VF.pdf

DISARM Gap Techniques

Two Techniques were introduced to plug gaps in the DISARM Framework; T0115.004: Send Message (a content delivery method that isn't 'posting'), and T0143.005: Compromised Persona (for use when an account has been compromised, but its previous identity is maintained, useful both for Fact Checks and across influence operations).

T0115: Post Content

Name	Description
T0115.004: Send Message	An actor delivered content by sending a message over direct or group messaging, or email.

T0143: Persona Legitimacy

Name	Description
T0143.005: Compromised Persona	This Sub-Technique is used to document actions taken by threat actors using a compromised asset, while maintaining the asset's pre-existing presented identity.

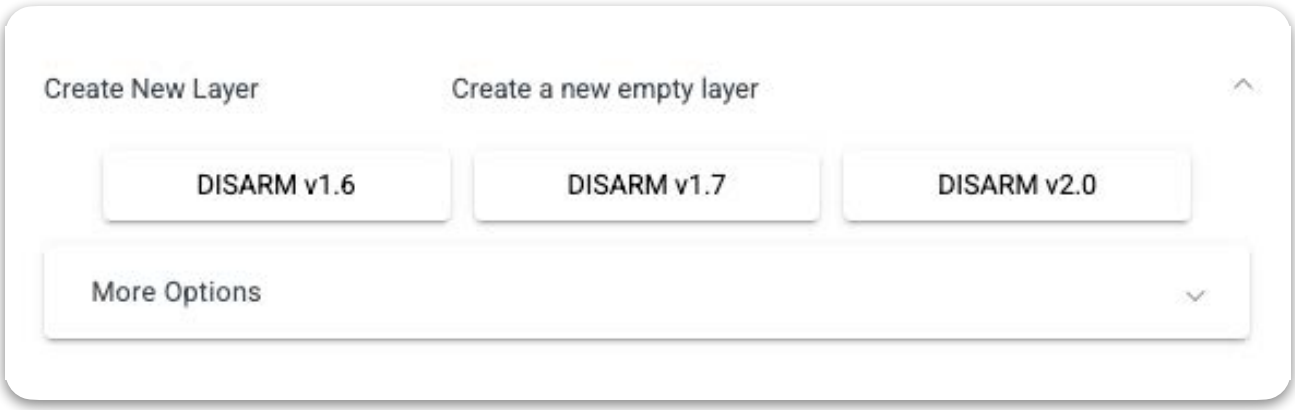
T0143.005: Compromised Persona

Source: https://www.theweek.in/news/biz-tech/2025/05/09/fact-check-did-pakistan-beg-for-money-from-world-bank-after-operation-sindoor.html

Multi Version Support

With this update we introduce multi-version support to the DISARM Navigator. When selecting "Create New Layer", you will be able to pick from previous, current,

and beta versions of the DISARM Framework. This change will make it easier for users to transition to new versions of the Framework.



Changes to Existing Content

This section describes changes made to existing items in the DISARM Framework.

Overview

Tactic, Technique, or Sub-Technique	Change Type
T0086.003: Deceptively Edit Images (Cheap Fakes) T0087.002: Deceptively Edit Video (Cheap Fakes) T0088.002: Deceptively Edit Audio (Cheap Fakes)	• Replaced by T0165: Edited Content
T0085.001: Develop AI-Generated Text T0086.002: Develop AI-Generated Images (Deepfakes) T0087.001: Develop AI-Generated Videos (Deepfakes) T0088.001 Develop AI-Generated Audio (Deepfakes)	• Replaced by T0166: AI-Generated Content • Replaced by T0166.001: Deepfake Impersonation
T0023: Distort Facts	• Replaced by T0162: Reframe Context
T0023.001: Reframe Context	• Changed to T0162: Reframe Context • Moved to TA14: Develop Narratives • Description Updated
T0023.002: Edit Open-Source Content	• Changed to T0169: Edit Open-Source Content • Moved to TA09: Deliver Content

T0016: Create Clickbait	• Changed to T0167.001: Use of Clickbait • Made a Sub-Technique of T0167: Issue with Content's Headline • Moved to TA06: Develop Content • Description Updated
TA08: Conduct Pump Priming	• Deprecated
T0044: Seed Distortions	• Replaced by T0068: Respond to Breaking News Event or Active Crisis • Replaced by T0049: Flood Information Space • Replaced by T0020: Trial Content • Replaced by T0135.003: Subvert
T0042: Seed Kernel of Truth	• Replaced by T0162: Reframe Context • Replaced by T0160.004: Information is Misleading • Replaced by T0160.001: Information is Verified
T0045: Use Fake Experts	• Replaced by T0097.108: Expert Persona • Replaced by T0143.002: Fabricated Persona
T0046: Use Search Engine Optimisation	• Moved to TA06: Develop Content
T0020: Trial Content	• Moved to TA09: Deliver Content

TA06: Develop Content

Deceptively Edit Image/Video/Audio (Cheap Fakes)

Deprecated

With the introduction of T0165: Edited Content, the following Techniques have been deprecated:

- T0086.003: Deceptively Edit Images (Cheap Fakes)
- T0087.002: Deceptively Edit Video (Cheap Fakes)
- T0088.002: Deceptively Edit Audio (Cheap Fakes)

There was too much overlap with these Techniques and T0165: Edited Content. "Cheap Fakes" are altered media made through 'conventional and affordable technology', which was a factor analysts have very little visibility over, so it was decided not to maintain the ability to make the assertion that an edit was made 'cheaply'.

To document which format the edited content has come in, analysts can use the format type alongside T0165: Edited Content. For example, an edited image can now be documented using both T0165: Edited Content and T0086: Develop Image-Based Content.

Develop AI-Generated Text/Image/Video/Audio Content (Deepfakes)

Deprecated

With the introduction of T0166: AI-Generated Content, the following Techniques have been deprecated:

- T0085.001: Develop AI-Generated Text
- T0086.002: Develop AI-Generated Images (Deepfakes)
- T0087.001: Develop AI-Generated Videos (Deepfakes)
- T0088.001 Develop AI-Generated Audio (Deepfakes)

There was too much overlap with these Techniques and T0166: AI-Generated Content and T0166.001: Deepfake Impersonation to keep them. These new Techniques allow for analysts to differentiate between AI-Generated content and Deepfakes, rather than using a single Technique for both.

To document which format the AI-Generated content has come in, analysts can use the format type alongside T0166: AI-Generated Content. For example, AI-Generated Text can now be documented using both T0166: AI-Generated Content and T0085: Develop Text-Based Content.

T0023: Distort Facts

Description: Change, twist, or exaggerate existing facts to construct a narrative that differs from reality. Examples: images and ideas can be distorted by being placed in an improper content

Deprecated

T0023: Distort Facts has significant overlap with T0162: Reframe Context, and as such it can be used in its stead.

T0023.001: Reframe Context

Description: Reframing context refers to removing an event from its surrounding context to distort its intended meaning. Rather than deny that an event occurred, reframing context frames an event in a manner that may lead the target audience to draw a different conclusion about its intentions.

Changed to T0162: Reframe Context and Moved to TA14: Develop Narratives

As T0023: Distort Facts was to be deprecated, T0023.001: Reframe Context needed a new home. It was changed to be its own Technique, with sub-techniques documenting common ways in which context is reframed. At this time it was also

moved to Develop Narratives, where it was better suited (the narrative a piece of content is presented with is what frames (or reframes) its context).
Description Updated

T0162: Reframe Context has had its description slightly updated to remove some of the judgemental or intentional language, in line with feedback that analysts sometimes felt uncomfortable asserting intent behind an observed behaviour.

Instead of stating that content has been removed from its original context with the intent of distorting its meaning, we now just state that it has been removed from its original context, which may reframe it in such a way that it reframes it.

T0023.002: Edit Open-Source Content

Description: An influence operation may edit open-source content, such as collaborative blogs or encyclopaedias, to promote its narratives on outlets with existing credibility and audiences. Editing open-source content may allow an operation to post content on platforms without dedicating resources to the creation and maintenance of its own assets.
Made T0169: Edit Open-Source Content and Moved to TA09: Deliver Content

As T0023.002: Edit Open-Source Content's parent Technique was being deprecated, it had to be moved or removed.

Because it could not reasonably be documented with existing Techniques, it was moved to be its own Technique under TA09: Deliver Content. This Tactic was chosen over TA06: Develop Content as threat actors are editing open source material as a method of delivering their content to a target audience.

TA05: Microtarget

T0016: Create Clickbait

Description: Create attention grabbing headlines (outrage, doubt, humour) required to drive traffic & engagement. This is a key asset.
Changed to T0167.001: Use of Clickbait

T0016: Create Clickbait has been changed to T0167.001: Use of Clickbait. This name change broadens the scope to not document the *creation* of clickbait, but its generic use.
Made a Sub-Technique of T0167: Issue with Content's Headline

T0167.001: Use of Clickbait has been made a Sub-Technique of the newly introduced T0167: Issue with Content's Headline. This allowed us to introduce non-clickbait issues with headlines, such as T0167.002: Title Misrepresents Content.
Description Updated

We've improved the description of T0167.001: Use of Clickbait to better explain what Clickbait is (basically, a headline which creates a knowledge gap instead of

functioning as a full story in itself - think “Analysts SHOCKED by What DISARM Did to T0016: Create Clickbait?!” vs “DISARM Changed T0016: Create Clickbait’s Description, Name, and Moved It”), and provide more examples of different types of Clickbait.

Moved to TA06: Develop Content

T0016: Create Clickbait used to be a Technique under the Tactic “Microtarget”, but this doesn’t make much sense as a Tactical goal of Clickbait, which broadly tries to get any viewer to view its content by creating enticing information gaps in headlines. It now lives under TA06: Develop Content.

TA08: Conduct Pump Priming

Description: Release content on a targeted small scale, prior to general release, including releasing seed. Used for preparation before broader release, and as message honing. Used for preparation before broader release, and as message honing.

Deprecated

This Tactic has been highly contentious, with people confused about what it means to Prime the Pump (we believe it is related to boating). As such, we are choosing to deprecate it. Conduct Pump Priming is described as trialling content before larger messaging, a behaviour which can be documented under T0020: Trial Content

T0044: Seed Distortions

Description: Try a wide variety of messages in the early hours surrounding an incident or event, to give a misleading account or impression.

Deprecated

This Technique’s description is confusing, its short description seems to cover the actor publishing a variety of different messages soon after an incident occurs both to “try” the messaging, and to give a misleading account.

So this Technique covers:

- Behaviour: Lots of messaging after breaking news
- Objective: Trialling narratives
- Objective: Misleading target audiences

Another issue with T0042: Seed Distortions is that its name doesn’t give a clear summary of its description - Seed Distortions doesn’t convey these three elements previously identified. As such, it may be the case that analysts have used Seed Distortions to document things other than what is described above. With feedback, we may need to introduce new techniques to address uncovered use cases.

For now, these elements can be documented using:

- Behaviour: Lots of messaging after breaking news
    - T0068: Respond to Breaking News Event or Active Crisis

    - T0049: Flood Information Space
- Objective: Trialling Narratives
    - T0020: Trial Content
- Objective: Misleading target audiences
    - T0135.003: Subvert

T0042: Seed Kernel of Truth

Description: Wrap lies or altered context/facts around truths. Influence campaigns pursue a variety of objectives with respect to target audiences, prominent among them: 1. undermine a narrative commonly referenced in the target audience; or 2. promote a narrative less common in the target audience, but preferred by the attacker. In both cases, the attacker is presented with a heavy lift. They must change the relative importance of various narratives in the interpretation of events, despite contrary tendencies. When messaging makes use of factual reporting to promote these adjustments in the narrative space, they are less likely to be dismissed out of hand; when messaging can juxtapose a (factual) truth about current affairs with the (abstract) truth explicated in these narratives, propagandists can undermine or promote them selectively. Context matters.

Deprecated

Seed Kernel of Truth was used to document cases where there were elements of truth in a campaign narrative. This terminology was confusing, particularly to non-native English speakers (what's a "Kernel" of Truth? How does one "Seed" it?).

DISARM has opted to deprecate Seed Kernel of Truth given the newly introduced Techniques which can more accurately document elements of what this Technique implied, such as:

- **T0162: Reframe Context:** True information has been taken out of its original context to frame it in a new way
- **T0160.004: Information is Misleading:** Information has been presented in a way which misleads viewers
- **T0160.001: Information is Verified:** True Information has been accurately presented

If it is the case that we receive feedback that the above Techniques are not enough of a replacement for T0046: Seed Kernel of Truth, we could keep this Technique with a new name (e.g. "Narrative Contains Element of Truth" or "Narrative Builds on Accurate Information"), and put it in TA14: Develop Narratives.

Otherwise, analysts may choose to continue using T0046: Seed Kernel of Truth - while it will no longer be displayed in future versions of the Framework, we still recognise completed reports which use T0046: Seed Kernel of Truth as having documented 'elements of truth in a narrative'.

T0045: Use Fake Experts

Description: Use the fake experts that were set up during Establish Legitimacy. Pseudo-experts are disposable assets that often appear once and then disappear. Give "credulity" to misinformation. Take advantage of credential bias

Deprecated

T0045: Use Fake Experts can be documented using previously introduced Techniques:

- **T0097.108: Expert Persona:** The asset is presenting as an expert
- **T0143.002: Fabricated Persona:** The asset's persona is fake

It may be the case that this Technique was used for things other than its description implied - for example that narratives cited fabricated experts, or that experts who had been discredited were cited. If we receive feedback that analysts used *Use Fake Experts* for these (or other) reasons, we should introduce new Techniques that help document them (alongside Techniques like T0163: Issues with Cited Academic Research and T0164: Issues with Presented Statistical Evidence). We will need to wait for feedback from existing users on its removal.

T0046: Use Search Engine Optimisation

Description: Manipulate content engagement metrics (ie: Reddit & Twitter) to influence/impact news search results (e.g. Google), also elevates RT & Sputnik headline into Google news alert emails. aka "Black-hat SEO"

Moved to TA06: Develop Content

Since we chose to deprecate TA08: Conduct Pump Priming, we needed to find a new home for T0046: Use Search Engine Optimisation. We moved it to TA06: Develop Content, as SEO must be considered when a content is developed.

T0020: Trial Content

Description: Iteratively test incident performance (messages, content etc), e.g. A/B test headline/content engagement metrics; website and/or funding campaign conversion rates

Moved to TA09: Deliver Content

While this Technique seems difficult to detect, and has a potentially misleading name when compared to its description, the decision was made to maintain the Technique until we were better able to address its potential issues, and understand how it was currently being used by analysts.

Until such time, TA09: Deliver Content is where it can reside, given the removal of TA08: Conduct Pump Priming.

Annex 15 - DISARM v2.0 Actions Rules

Actions Rules

This document describes the Rules that apply to different Observations in the Action section of the DISARM v2 Framework. Observations are individual things you can describe about a potential incident. Action are a type of Observations specifically focused on describing the - actions - that are taken by an Asset.

Rules for Procedures in DISARM v2

DISARM v2 makes allows you to make standardised Procedures by associating different Observations to better describe a potential incident. It's useful to be able to associate Observations because it allows analysts to tell a better story about what they've seen happening in a potential incident, to describe the scene, and lay an evidence-based groundwork for assessments of technique usage, motive, and attribution.

This section provides general information about how to make a Procedure in DISARM v2.

Rule - Elements of a Procedure

A Procedure documents an Asset taking some kind of Action on Material or another Asset.

An Asset is a tool used to take actions online; things like accounts, websites, and other digital tools. An example of an Observation which documents an Asset is Account; i.e. “I observe that the asset being used in this incident is an account that exists on a platform”.

An Action is something an Asset does (like posting, sending a message, or adding another account). An example of an Observation which documents an Action is Send; i.e. “I observe that the action being taken by the asset is sending something privately to an individual or group”.

Material is things like content, posts, or merchandise. An example of an Observation which documents Material is Message; i.e. “I observe that the thing being acted upon is in the format of short form material sent privately to an individual or group”

Procedures are made up of up to three elements. The first slot in a Procedure identifies what type of Asset is taking Action. The second slot identifies the Action it was taking. The third slot identifies the Material or Asset the Action was taken on.

Rule - Procedure Makeup

To make a procedure, open a pair of brackets to indicate you’re starting a procedure. To move through each element, use a greater-than sign.

A Procedure which documents an asset taking action on material would look like:

(Asset > Action > Material).

A Procedure which documents an asset taking action on another asset would look like:

(Asset > Action > Asset).

You use the “>” symbol to denote you’re moving to the next section of the Procedure

For example, an account which takes the action of sending a message would be documented using:

(Account > Send > Message)

This is called the “Top Level” of a Procedure. Each Observation has Rules which identify whether they can be used in the Top Level of a Procedure.

Rule - Using Observations to describe other Observations in a Procedure

Observations in a Procedure can be described by other Observations. This is useful to provide more detail about different sections of the Procedure.

To do this, open a new pair of brackets after the Observation you want to describe, i.e.

(Observed Observation (Observation which describes it))

For example, Compromised Asset is an Observation which can be used to describe different types of Asset. The following Procedure can be used to document a compromised account:

(Account (Compromised Asset))

You can describe an Observation with multiple Observations within the same level of a Procedure. To do this, list all relevant Observations separated by a comma.

For example, a compromised account presenting as being controlled by a nonexistent entity would look like:

(Account (Compromised Asset, Asset Presents as Being Controlled by Nonexistent Entity))

Each Observation has Rules which identify whether they can be used to describe another Observation, and if so, which Observations they are able to describe.

Rule - What Constitutes a Procedure

Procedures don't have to have all three sections at the Top Level (i.e. Asset > Action > Material / Asset). Procedures are:

1) One or more Observations in a bracket at the top level, paired with ">" - like this:

   (Asset > Action)

   A Procedure with only two items at the top level can be one of:

   - An asset taking an action: (Asset > Action)
   - An action being taken on a material: (Action > Material)
   - An action being taken on an asset: (Action > Asset)

   You **cannot** have (Asset > Material)

2) A Procedure can also be an Observation of any type which has been described by another Observation - like this:

(Account (Compromised Asset))

3) A Procedure can be any combination of (1) and (2) - like this

(Account (Compromised Asset), Post)

Rules for Action

This section of the document provides Rules which apply to all Observations in the Action section of the framework.

These rules exist to help reduce some of the complexity around using the framework (even if it might not seem like that's true when you're half way through reading the document). They help take the framework from a list of hundreds of different Observations, to a filtered list of relevant Observations.

Rules are made up of two sections; Rule Text, and Rule Explanation. Rule Text provides the exact text that is in effect when a Rule is applied to an Observation. Rule Explanation provides more information about how the Rule works.

Rule - Modular Rules

Rule Text

This Rule applies to all Rules.

If a Rule's Rule Text ends in a colon, then the Rule is completed by one or more Observations.

The Observations which complete the Rule are specified within the Rule section of an Observation.

Rule Explanation

Rule - Described by is a rule which functions to limit what Observations can be used to describe another Observation the Rule applies to. Its Rule Text is:

> The Observation this Rule applies to can be described by the following Observations:

The Rule Text ends in a colon, so it is completed within each instance of the Rule's application by listing the relevant Observations.

As an example, this rule could be applied to Message (which is a Material Observation) in the following way:

Rule - Described by:

- Text Content
- Image Content
- Video Content

This would mean the only Observations which could be used to describe Message are Text Content, Image Content, and Video Content, e.g.:

(Message (Text Content))

If Rule - Modular Rules is completed by an Observation which contains other Observations, then all Observations contained by it complete that Observation.

Rule - Usable in Top Level

Rule Text

This Observation can be used in the Top Level of a Procedure.

Rule Explanation

This Rule helps clarify whether an Observation functions as a Top Level item in a Procedure.

For example, Send has this rule applied to it. Send can therefore be used in the Top Level of a Procedure.

“Platform Fails to Disclose Use of Automation” does not have this rule applied to it. Instead it has the rule Rule - Descriptor applied to it. This means it is an Observation which exists only to describe other Action Observations.

“Platform Moderation Action” has neither “Rule - Usable in Top Level” or Rule - Descriptor applied to it. This means it is an Observation which exists only to categorise other Observations in the framework, and cannot be applied in any way as part of a Procedure.

Rule - Observation Type

Rule Text

Each Observation are one of Asset, Action, or Material. This Observation is the following type:

Rule Explanation

The following Rule is applied to all Observations in the Action category

Rule - Observation Type: Action

This functions to differentiate between Asset, Action, and Material in the Observations framework.

Rule - Online or Offline

Rule Text

Each Observations are one of Online or Offline. This Observation is the following type:

Rule Explanation

This Rule helps differentiate between Online and Offline actions. 99% of Observations that exist in the framework have the following Rule applied to them:

Rule - Online or Offline: Online

This lays the groundwork for documenting offline activity at a later date.

Observation Filtering Rules

These Rules help filter which Observations are relevant based on the Observation you chose

Rule - Described by

Rule Text

This Observation can be described by the following Observations:

Rule Explanation

Rule - Described by helps filter the framework based on what Observations has been selected.

As an example, this rule could be applied to Message (which is a Material Observation) in the following way:

Rule - Described by:

- Text Content
- Image Content

- Video Content

This would mean the only Observations which could be used to describe Message are Text Content, Image Content, and Video Content, e.g.:

(Message (Text Content))

Rule - Lead to

Rule Text

This Observation can lead to the following Observations:

Rule Explanation

Rule - Lead to filters what other Observations can come **after** the Observation this Rule is applied to is used in a Procedure (i.e. what Material or Asset can appear after a ">" when the Action this Rule applies to is used).

For example, Send has the following rule:

Rule - Lead to:

- Message
- Email

This means that if Send is the Action in a Procedure, the **only** Observations it can lead to are Message or Email, e.g.:

(Send > Message)

Rule - Lead to Me

Rule Text

This Observation can only be used if it is preceded by the following Observations:

Rule Explanation

Rule - Lead to Me filters what Observations can come **before** the Observation this Rule is applied to is used in a Procedure (i.e. what Asset can appear before a ">" when the Action this Rule applies to is used).

For example, Deliver has the following rule:

Rule - Lead to Me: Platform

This means that if Deliver is the Action in a Procedure, the **only** Observation which can lead to it is Platform, e.g.:

(Platform > Deliver)

Note, we don't need both Rule - Lead to and Rule - Lead to Me once the framework is completed (for example, you can imagine Platform having Rule - Lead to: Deliver, making this rule redundant) - but while different parts are being developed, we will have both Rule - Lead to Me and Rule - Lead to, and standardise it into (probably) just Rule - Lead to.

Descriptor Rules

These Rules only apply to Observations used as descriptors.

Rule - Descriptor

Rule Text

This Observation can be used to describe the following Observations:

Rule Explanation

This Rule helps clarify whether an Observation functions as a descriptor in a Procedure.

"Platform Fails to Disclose Use of Automation" has this Rule applied to it. This means it is an Observation can be used as a descriptor of Action Observations.

Send does not have this Rule applied to it. Instead, it has the Rule Rule - Usable in Top Level applied to it. This means it is an Observation which exists only to document the Action taken by an Asset at the Top Level of a Procedure

"Platform Moderation Action" has neither "Rule - Usable in Top Level" or Rule - Descriptor applied to it. This means it is an Observation which exists only to categorise other Observations in the framework, and cannot be applied in any way as part of a Procedure.

Rule - Mutually Exclusive Descriptor

Rule Text

This Observation is a descriptor which cannot be used as a descriptor alongside the following Observations:

Rule Explanation

This Rule has not been used in the Action category. It has been used in the Asset category.

The Asset Observations "Newly Created Asset" has the following rule:

> Rule - Mutually Exclusive Descriptor: Pre-Existing Asset

This would mean you can't describe an asset as both "Newly Created" and "Pre-Existing".

Rare Rules

These are advanced Rules which apply to very few Observations.

Rule - Action Action

Rule Text

Using this Action changes Rule - What Constitutes a Procedure to allow another Action to follow this Action in the same level of the Procedure.

This Action Lead to the following Action Observations:

Rule Explanation

An Action with the Rule - Action Action may look like:

> (Asset > Action > Action > Material/Asset)

This is done to enable modular actions, or actions leading to other actions. For example, Suggest Action is an Action available to a Platform in which it prompts a user to take a given action. A platform may take the action of suggesting to a user that they take the action of joining a group, which would look like:

> (Asset > Suggest Action > Join Community > T0151.002: Online Community Group)

Rule - Conditional Rule

Rule Text

Apply the next Rule only if this condition is met:

Rule Explanation

Some Rules only apply in certain situations. This Rule lays out those situations.

For example, Deliver is an Action which can lead to Search Results or Advertisement. Different Rules apply based on which type follows Deliver.

So Deliver could have the following rule:

> Rule - Conditional Rule: If Deliver Lead to Search Results
>
> Rule - Described by: Search

Which would mean if Deliver is used to document the action of delivering Search Results, then Deliver can be described by the Search which was submitted to it. For example:

> (Platform > Deliver (Search (Text Content (Coded Terminology, CSAM))) > Search Results (Image Content (CSAM)))

Rule - Described by Procedure

Rule Text

This Observation can be described by a Procedure which contain at least one of the following Observations:

Rule Explanation

Typically Observations can only be described by a single Observation, or a collection of unassociated Observations. For example, a Post could be described by both Text Content and Image Content, indicating the post contained image and text. It would look like:

> (Post (Image Content, Text Content)).

This rule denotes a case when a Observations can be described by a Procedure, which helps document Action which necessitates documentation of other Procedures.

For example, the Comment Action has the following rule:

> Rule - Described by Procedure: [Publication Type]

This is to enable documenting not only the post that was commented on, but information about the Asset that posted it.

Continuing the example, a fake account commenting an image on the post of a real fact checker would look like:

([Account] (Fake Identity) > [Comment] ([Account] (Fact Checker (Real Identity)) > Publish > Post) > [Post] ([Image Content]))

[Rule - Apply Rule to Following Observation]

Rule Text

Apply the next Rule to [Observations] which follow this [Observation] in the [Procedure]:

Rule Explanation

[Rule - Apply Rule to Following Observation] changes any Rules which apply to the [Observation] which follows the [Observation] this rule applies to in a [Procedure].

This has been used to apply [Rule - Described by Procedure] to [Observation] which follow on from an [Action] that necessitate larger descriptions of what has been acted upon.

For example, the moderation action "[Platform] Removes [Publication] from [Platform]" has the following rules:

> [Rule - Lead to]: [[Publication] Type]
>
> [Rule - Apply Rule to Following Observation]
>
> [Rule - Described by Procedure]: [[Publication] Type]

This combination means:

1. "[Platform] Removes [Publication] from [Platform]" must be followed by an [Observation] from the "[Publication] Type" category.
2. "[Platform] Removes [Publication] from [Platform]" applies the replaces the Rules of the next [Observation] in the [Procedure] with the following rule:
3. This [Observation] is described by a [Procedure] which contains an [Observation] from the "[Publication] Type" category.

This combination of Rules means that when a Platform removes a Publication from the platform, context about how the Publication got there can be documented.

Imagine WhatsApp removed a Message from their platform posted by a fake account. This could be documented using:

> (Chat Platform > Platform Removes Publication from Platform > Message (Account (Fake Account) > Send > Message)

Table of Rare Rule Implementation

This table details how some of the “Rare Rules” have been implemented:

Special - Comment on Existing Public...
 	
Applies to: **Comment**
	Rule - Described by Procedure: [Publication Type]
	
	--- 	
	
	This is because comments are left on existing Publication, rather than being new Publication made to an Online Asset.
	
	The whole Procedure associated with the Publication which is commented on can Describe the Comment Action.
	
	--- 	
	
	For example:
	
	An account with a fake identity commenting on another account with a fake identity’s post which contains an image depicting fake graffiti with a post tagging a real fact checker and asking them to produce a fact check
	
	- Account
	  - Fake
	- Comment
	  - Account
	  - Fake
	  - Publish
	  - Post
	  - Image
	  - Fake Graffiti
	- Post
	  - Text
	  - Tag
	  - Account
	  - Fact Checker
	  - Real
	  - Solicit Action
	  - Produce
	  - Post
	  - Fact Check
	
	An account commenting on another account’s image post with racist abuse
	
	- Account
	- Comment
	  - Account
	  - Publish
	  - Post
	  - Image

	- Post
	  - Text
	  - Abusive Content
	  - Discriminatory Content

Special - Amplifying Existing Publicati...
 	
Applies to: **Amplify** **Online Publication**
	Rule - Apply Rule to Following Observation:
	
	Rule - Described by Procedure: [Publication Type]
	
	--- 	
	
	This is to fully document the Publication which has been Amplified by the Asset taking the amplification action.
	
	--- 	
	
	For example:
	
	An account with a fake identity reposting a post made by another account with a fake identity which contains an image depicting fake graffiti
	
	- Account
	  - Fake
	- Repost
	- Post
	  - Account
	  - Fake
	  - Publish
	  - Post
	  - Image
	  - Fake Graffiti
	
	A fake account emote reacting to a post by another fake account which contained an image of fake graffiti
	
	- Account
	  - Fake
	- Emote React
	- Post
	  - Account
	  - Fake
	  - Publish
	  - Post
	  - Image
	  - Fake Graffiti
	
	A fake account which forwards a Bullshit message to a group chat which had been DM'd do them by a fake account
	
	- Account
	  - Fake

	- Forward
	  - Group Chat
	- Message
	  - Account
	  - Fake
	  - Send
	  - Account
	  - Fake
	  - Message
	  - Text
	  - Bullshit

Special - Moderating Existing Publicat...
 	
Applies to:
 	
• Platform Takes Moderation Action
• Platform Denies Infringement is Infringing
	Rule - Conditional Rule: If this Observations Lead to Publication
	
	Rule - Apply Rule to Following Observation:
	
	Rule - Described by Procedure: [Publication Type]
	
	--- 	
	
	This is to fully document the Publication which has been Moderated by the Platform taking the Moderation Action.
	
	--- 	
	
	E.g.
	
	- Platform
	- Remove Publication from Platform
	- Post
	  - Account
	  - Fake
	  - Publish
	  - Post
	  - Non Infringing Content

Special - Platform Deliver Search Res... Applies to: • Deliver • Platform Suppresses Search Results	Rule - Conditional Rule: If this Observations Lead to Search Results Rule - Described by Procedure: Search --- This is to document what was searched, and who searched it. --- E.g. - Platform - Deliver

	- Account
	  - Young Person
	 - Submit
	 - Search
	  - Text
	  - Eating Disorder Content
	  - Coded Terminology
	- Search Results
	 - Image
	  - Eating Disorder Content
	  - Content Goes Against Platform Policy

Special - Report Interaction Applies to: • Submit • Platform Fails to Act on Report in Timely Manner	Rule - Conditional Rule: If this Observations Lead to Report Rule - Apply Rule to Following Observation: Rule - Described by Procedure: • User • Community • [Content Type] • Post • Message • Call • File • Stream • Email This is to document what was reported

Annex 16 - DISARM v2.0 Assessments Framework

Purpose

This roadmap is intended to guide analysts in incorporating DISARM observations as the foundation for producing structured threat assessments. The goal is to support informed response decisions by clearly outlining how observable behaviors, assets, and content can be transformed into evidence-based threat judgments. Ideally, this process can be refined into a shared methodology that is replicable across government teams and other relevant organizations.

Defining a Threat Assessment

A threat assessment is an analyst's judgement evaluating and assessing the potential risk an observed behavior or pattern presents to a specific population, institution, system, or audience, and the ways in which it does so.

Threat assessments serve multiple purposes: they catalogue emerging threats before they escalate at key moments, support decision-making around response options, and provide a basis for scenario testing based on assessed risk. **Over time, assessments can function as case studies — supporting knowledge sharing via ISACs, training new analysts on recognition patterns, and helping identify or hypothesize links between isolated incidents and broader, multi-vector campaigns.**

DISARM observables help supply the behavioral characteristics that form the foundation of an assessment. However, producing a full assessment requires analysts to interpret and structure that evidence, often bringing in supplemental contextual, geopolitical, or narrative information.

Observations found within a threat assessment: What Disarm provides vs additional info provided by Analysts

Orange = CFraming

Assets

- **Provided by DISARM**: Accounts, content objects, communication channels, and types of platforms used.

- **Analyst adds**: Geographic and platform metadata, prior asset history (e.g. previous flags or uses), potential linkages to known campaigns, ownership inference where possible.

Content Type

- **Provided by DISARM**: Format tagging (image, video, text, synthetic media), general function.

- **Analyst adds**: Framing, emotional tone, ideological function, narrative lineage (e.g. recycled or original), deception techniques, and intent clarity.

Actions

- **Provided by DISARM**: Full list of observable tactics and techniques (TTPs), tagged in context.

- **Analyst adds**: Severity (e.g. scale, escalation), intent (reactive vs planned), signs of coordination, and campaign context or phase.

Audience

- **Provided by DISARM**: Not directly specified, though some tagging hints at intended target types.

- **Analyst adds**: Target demographics, community characteristics, ideological alignment, risk level, platform segmentation.
  (Could be formalized with decision-tree mockups or interface tags in the future?)

Timeline

- **Provided by DISARM**: Implicit timing via behavior logs, asset timestamps, and content activity (roadmap tagging potential).

- **Analyst adds**: Full event alignment (e.g. elections, protests), temporal clustering across platforms, long-term amplification tracking, and campaign phasing.

Harms

- **Provided by DISARM**: Partial coverage via specific playbooks (e.g. TFGBV, impersonation, suppression).

- **Analyst adds**: Broader categories of civil, political, institutional, or reputational harm — including financial, psychological, physical, or privacy-related risk.
  (Smart to update and expand harms on playbooks like the TFGBV to include further common. Intersectional and real-world risks.)

Additional Potential Analyst Contributions (Cross-Cutting):

- Actor attribution (based on behavior, infrastructure, and narrative alignment)

- Coordination patterns and tactical escalation

- Platform and institutional response analysis

- Confidence scoring or uncertainty notes between network connections and for attributions (useful for ISAC-style sharing)

Assessment Categories (Open to add/ modify onto this list)

These categories allow analysts to structure their conclusions in a way that supports decision-making and pattern recognition across campaigns while also allowing for risk prioritization according to a sliding scale of harm. Each category draws on DISARM observables (where applicable), but also requires analyst judgment, contextual intelligence, or cross-platform insight (the last two bullet points especially as well as the kill chain stage evaluation).

This structure helps enable consistent, reusable assessments that can scale across teams, tools, and information-sharing environments:

- **Threat Type**
- **Actor Type**
- **Target Audience + Campaign Target**
- **Motive and Strategic Objective**
- **Timing (Event Alignment with other major current events)**

- **Infrastructure and Capabilities**
- **Breakout Scale / Kill Chain Stage**
    - How far has this activity progressed? Using a scale (e.g. Stage 1–6), analysts can assess: Seeding → Amplification → Targeting → Mainstream pickup → Engagement → Real-world consequences. This can help determine the urgency of response. Some threats are low-traction and are best simply tabbed and catalogued, others are already escalating and deserve a real response.
    - Vic: We can explore a variety of "kill chains" now we've un(kill)chained ourselves from the old organising principle.
        - Adam: We could have a column (like the old Tactics?) which is "Kill Chains" (or whatever) and give people options for which KC they want?
- **Vulnerabilities Exploited**
- **Harms (Real and Potential)**
- **Content Authenticity**
- **Content Reliability**

Example: Election Interference Threat Assessment including Portal Kombat style elements

*I can expand the analytical thinking process for every step. I was unsure, at this stage, how much detail you were looking for.

Mock Simulation Scenario: A video falsely claiming an investigation is being conducted into how mail-in ballots are being used to interfere with official election results is hosted on a spoofed version of a national news website (the spoof has a national ccTLD) , and amplified by Telegram channels, X accounts and specific Facebook groups containing amateur investigators and right wing political conspiracy. The link for the news website appears legitimate to casual scrollers. The video is structured like a news report with subtitled language.

Using the above categories, how can a NATO analyst conduct a threat assessment and arrive at a conclusion of perceived harm and risk using DISARM 2.0 classifications?

The below walk-through example uses tags from both the 2.0 Prototype Victoria shared with me and the traditional framework.

1. Threat Type

What kind of threat is this?

- Begin by identifying **observable behaviors**: hosting deceptive content on a cloned site, sharing it in political conspiracy groups and identifying and analyzing the nature of the content itself

- Use this to tag DISARM Assets and Content like:
- ASI4.4 Lookalike Domain *(Domain names designed to appear similar to legit)*
- CN3 Content Presentation
    - CN3.5 Content Presented with False Context
    - CN3.7 Content Presented as Depicting Current Events
    - CN3.8 Fabricated Content Presented as Real
    - CN3.9 Source Embedded in Media
- CD3.5 Subtitled Speech
- CA3.7 Context Removed from Content
- CD5 Content Anomalies
    - CD5.1 Fabricated Content
- CD6.1 News Report

Judgment call: this appears to be an **informational threat**, elevated by infrastructure manipulation.

2. Actor Type

Who might be behind this?

- Attribution is initially unclear, but the use of a spoofed ccTLD site gives a lead on next steps. raises suspicion.

- Analysts can conduct WHOIS **domain lookups**, checking registration history, DNS records, and shared infrastructure identifiers (such as registrant email overlaps, or hosting patterns).

Tools involved in this step (inferred from scenario in VIGINUM Report):

- WHOIS records
- DNS history tools (e.g. RiskIQ, DomainTools, PassiveTotal)
- Reverse IP lookups
- Registrant metadata comparison (emails, names, name servers)
- Prior campaign tagging and open repositories (e.g. Graphika reports, VIGINUM datasets)
- **IF** cross-referencing this data against existing OSINT repositories and prior assessments reveals:
    - The spoofed site shares **technical infrastructure characteristics** with previously identified fake portals attributed to the **Portal Kombat network**
    - The domains in question were registered via a **Crimea-based web services company**, already flagged in VIGINUM's reports as a consistent infrastructure provider for Russian-aligned influence activity
- **THEN** based on this, the activity is assessed as part of a **coordinated network**, likely a **foreign non-state proxy or state-aligned actor** using outsourced technical infrastructure.
- T0097.202 News Outlet Persona
- Tactic: TA16 Establish Legitimacy

3 and 4 : Target Audience & Campaign Target + Strategic Objective / Motive

Who's being influenced? Who's being attacked or undermined? + *What's the campaign trying to achieve? Is this to provoke unrest? Suppress voter turnout? Undermine results?*

- Analysts can search through the comments for opinions/ comments/ subtweets that have gained a lot of traction or have been pinned (Youtube, Facebook, Twitter) for overall opinions and viewpoints being amplified.
- The **audience** is likely right-leaning or undecided voters prone to institutional distrust

- The **target** is trusted national mainstream sources of information + overall democratic legitimacy

- DISARM tag: TA4 – Undermining civic trust
- Use narrative framing + platform choice + timing to infer motive

- Possible tag from campaign development: CD3.5 – Delegitimize election infrastructure

5. Timing / Event Alignment

Why now?

- Correlates with voting dates, controversial election campaigning or breaking political news

- Given the way the timing of current events align with the posting patterns of these videos: DISARM behavior tag: T0068 Respond to Breaking News Event or Active Crisis .// 2.0 Prototype: CN2 Respond to Breaking News Event or Active Crisis

6. Infrastructure / Capabilities

How advanced is this?

- Hosting on a ccTLD shows planning and a certain amount of technical knowledge/ resources.

- Combined with coordinated cross-platform spread, and a presence across multiple social media channels, it's not a spontaneous effort. Required a certain degree of

planning and coordination.

- DISARM Tags that reflect this:
    - PAA6 Cross-Posting
    - PAA7 Networked Action
        - PAA7.2 Concurrent Networked Action
- **Resulting Assessment:** Moderate capability with potential towards high capability as investigation + perceived harms/ impact continues to be assessed.

7. Breakout Scale / Kill Chain Stage

How far has it spread? (First part of impact analysis)

- Initial seeding → Telegram groups → X influencers → Facebook clusters

- Catalogue posting patterns and use platform analytics, engagement volume, quote tweets to conduct initial impact assessment on how much this information has been disseminated and its outreach potential. DISARM Tags that reflect these observations:

T0094 Infiltrate Existing Networks

T0100 Co-Opt Trusted Sources

T0102 Leverage Echo Chambers/ Filter Bubbles

- **Tentative assessment at this stage:** A targeted amplification that is not yet mainstream but has successfully reached the fringes of society that it targeted.

8. Vulnerabilities Exploited

What public fears or tensions is this tapping into?

- Distrust of vote-by-mail

- Belief in media collusion / cover-up

- Cross-reference ongoing narratives in prior campaigns for integrated narratives (anti-elite sentiment, diaspora resentment, xenophobia etc.)

9. Harms (Real and Potential)

What tangible harms can we assess as likely consequences?

- Short-term: Confusion, misinformation spread, distrust and disillusionment on voting process

- Long-term should this gain further traction : Decreased voter confidence (and turnout?), baseless calls for recounts or audits, possible potential unstable protest coordination (see American example at the worst level case scenario). Overall: a lot of domestic distraction and damage control, leaving other areas open to vulnerability,

- Consider: risk of platform inaction or mislabeling

  Disarm Tags to flag that correspond with these potential harms and vulnerabilities (8 and 9)

- T0022.001 Amplify Existing Conspiracy Theory Narratives
- T0024 Distort Facts
- T0059 Play the Long Game

10. Content Authenticity / Reliability

Is the content fabricated, manipulated, or real?

- Video: uses likely real footage as well as some well known universal images with false narration or context. No initial signs of AI for this batch. (flag it as future potential for these kinds of videos).

- Hosted on a **spoofed news site.**

- Analyst labels it as **verifiably false, deceptively presented. DISARM 2.0 Tags/ Original Tags:**
- T0152.004 Website Asset
- T0097.202: News Outlet Persona

- T0143.003 Impersonated Persona

- CA5.4 Stock Media Content
- CA5.7 Content Previously Published Online

Reasonable Analyst Conclusion and Response Recommendations :

Based on the above assessment, it can be reasonably concluded that this is a targeted informational threat effort by coordinated Russian foreign adversaries to undermine trust in the electoral process, local media institutions, and broader democratic stability. The infrastructure and tactics align with previously observed behaviors associated with the **Portal Kombat network**, or, a closely affiliated actor or copycat operating within the Russian FIMI space. Attribution is made with **medium confidence** (using Graphika's scale), based on shared infrastructure characteristics and amplification patterns. The full network of threat actors (actual individuals, the extent of State involvement or related organizations affiliated with this online activity remains uncertain at this stage- analysts scouring previously known portal kombat links and persons of interests for further leads.)

Response recommendations for now:

- Definitely flag it and keep it on record.
- Share assessment summary across trusted partners and alert media channels targeted so they can issue clarifications of authenticity in their own right.
- Monitor for growing traction and the potential need for a cross-platform debunk.
- In the meantime, continue outreach to educate voters on the process and begin monitoring for related coordinated offline activity.

This modelled example could be used to support cross-team training, or shared threat libraries. (Video tutorial style walkthrough example with Julian)

Additional notes to expand on:

➢ DISARM BLUE is out of scope but a template for a sliding scale response framework similar to the ones we prepped in the workshops at Wilton Park can be useful in the future: Low-risk = monitor only, medium-risk = platform nudge, high-risk = coordinated debunk. With the type of debunk varying according to severity)

➢ **A Scenario Sketch for stress-testing the Assessments Framework Using the TFGBV Playbook: SCENARIO- A TFGBV attack targeting a Georgian journalist**

(from likely Russian actors) using AI generated images and impersonation with the aim of reputational suppression. The structure would follow a similar decision making process (likely with more of an emphasis on real life harms given the context throughout the thought process).

Literature/ Examples that helped shape this:

2023 EEAS Threat Report:
https://euvsdisinfo.eu/uploads/2023/02/EEAS-ThreatReport-February2023-02.pdf

Viginum Report on Portal Kombat:
https://www.sgdsn.gouv.fr/files/files/Publications/20240214_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_PART2_ENG_VF.pdf

Beyond Operation Doppelgänger: A Capability Assessment of the Social Design Agency:
https://mpf.se/psychological-defence-agency/publications/archive/2025-05-15-beyond-operation-doppelganger-a-capability-assessment-of-the-social-design-agency

Full List of DISARM 2.0 Prototype Tags used in walk-through example

ASSETS

DPC5- Digital Community Hosting Asset

- DPC5.2 Social Media Platform
- DPC5.4 Online Community Page
- DPC5.8 Chat Broadcast Group

AD6 Asset Origin- Information about where an asset originated from.

- ASI4 Online Infrastructure - Techniques related to online infrastructure components
- ASI4.4 Lookalike Domain- Domain names designed to appear similar to legit. domains.

ACTIONS

PAA4 Publish Content

PAA4.13 Upload File

PAA5 Amplify Content

PAA5.2 Repost Post

PAA5.4 React to Post

PAA5.5 Comment on Post

PAA6 Cross-Posting

PAA7 Networked Action

PAA7.2 Concurrent Networked Action

CONTENT

CN2 Respond to Breaking News Event or Active Crisis

CN3 Content Presentation

CN3.5 Content Presented with False Context

CN3.7 Content Presented as Depicting Current Events

CN3.8 Fabricated Content Presented as Real

CN3.9 Source Embedded in Media

CD3 Video Content

CD3.5 Subtitled Speech

CA3 Edited Content

CA3.7 Context Removed from Content

CD5 Content Anomalies

CD5.1 Fabricated Content

CD6 Content Style

CD6.1 News Report

CA5 Content Origin

CA5.4 Stock Media Content

CA5.7 Content Previously Published Online

Original DISARM Tags included:

T0022.001 Amplify Existing Conspiracy Theory Narratives

T0024 Distort Facts

T0059 Play the Long Game

T0068 Respond to Breaking News Event or Active Crisis

T0094 Infiltrate Existing Networks

T0096 Leverage Content Farms: T0149.003 Lookalike Domain

T0100 Co-Opt Trusted Sources

T0102 Leverage Echo Chambers/ Filter Bubbles

T0152.004 Website Asset

T0097.202: News Outlet Persona

T0143.003 Impersonated Persona

31.03
2026

ADAC.iO
Attribution Data Analysis
Countermeasures Interoperability

ADAC.io Publication

Date of Publication: 31.03.2026
Contact: victoria.smith@disarm.foundation
See more at adacio.eu

Funded by the European Union Horizon Europe Research and Innovation Program and the UKRI under the UK government's Horizon Europe funding guarantee. Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union, the Horizon Europe Research and Innovation Program or UKRI. Neither the European Union, the Horizon Europe Research and Innovation Program, nor the UKRI can be held responsible for them.

Co-funded by
the European Union