Government Communications

Resist

RESIST 3

A framework for building resilience to information threats.

communications.gov.uk

The RESIST 3 framework, developed by Dr James Pamment, Director of the Lund University Psychological Defence Research Institute, offers a structured approach for government communicators to build resilience against information threats. It guides users through recognising, monitoring, and analysing manipulated information, and then outlines strategic communication responses to counter its impact.

Contents

Foreword: Empowering communicators to build resilience

We live in an increasingly volatile world where the threats we face are both physical and online. Disinformation poses a global threat to our democratic societies, compromising national security, public safety and the integrity of information. A localised incident can drive civil unrest and escalate into a national emergency within days or hours. It is the top priority of any government to keep its citizens safe.

RESIST 3 covers the evolving nature of information threats, which blur the lines between misinformation and disinformation, whether they are individual activities or part of coordinated campaigns. It provides a systematic, evidence-based approach to help organisations and governments build and maintain societal and individual resilience.

Just like the lighthouse cutting through a storm on the cover, this updated RESIST framework illuminates the path to clear communication, guiding us through the chaos of information threats.

This latest version is a step forward from the 2021 RESIST 2 in three main areas:

- First, this framework has an enhanced focus on strengthening societal resilience, for which strategic communication is an important tool for building trust, responding to harmful content, and better equipping the public to withstand information threats.

- Second, we acknowledge how emerging technologies such as artificial intelligence (AI)-generated content and bot-driven influence campaigns can spread false narratives at an unprecedented speed and scale. Yet technology is also the most powerful tool in identifying and countering these threats. Communications responses need to harness these new technologies, and be smart, strategic and audience-focused.

- Third, we have developed our existing guidance for assessing threats and vulnerabilities, to better reflect the importance of knowing our own strengths, weaknesses, and priorities, before countering information threats. Vulnerability assessment is therefore a key step in understanding how likely a threat actor's behaviour is to cause problems to society.

We know that malign actors attempt to identify and exploit critical gaps in the response protocols of our democracies. But if you are willing to move quickly, test new approaches and determine the right communications responses, your organisation or government will emerge more resilient and better prepared to face future challenges.

Simon Baugh
Chief Executive,
Government Communications

Introduction: Building resilience to information threats and vulnerabilities

Originally developed in 2018, the aim of RESIST is to support communicators in reducing the impact of manipulated, false, and misleading information on wider society and national interests, in line with democratic values.

The framework is fundamentally about building resilience to information threats and improving the communications response. It promotes a consistent approach to the threat of mis- and disinformation by providing six steps to follow:

1. **R**ecognise mis- and disinformation
2. **E**arly warning
3. **S**ituational insight
4. **I**mpact analysis
5. **S**trategic communication
6. **T**racking effectiveness

Building resilience is the ability to withstand, adapt to, and recover from adversity. It is a key priority if we are to create continuity despite ongoing challenges. It enables individuals, communities and societies to maintain stability and function effectively in the face of disruption by proactively strengthening systems, structures, and individuals to mitigate risks and enhance long-term stability.

This involves improving internal capabilities to resist and recover from external pressures, reducing vulnerabilities, and demonstrating resolve, making societies less attractive targets and better equipped to face unanticipated threats. Government communications play a key role in building resilience by fostering long-term trust, developing contingency plans and capabilities in the medium-term, and effectively handling crises when they arise.

It is important to note that resilience-building is a collective effort that involves governments, the private sector, civil society, and individuals. It requires a whole-of-society approach, where multiple stakeholders collaborate to create stability. Although outside of the remit of RESIST, additional resilience-building activities, such as improving media literacy, making digital infrastructure more secure, and reinforcing democratic institutions, are also essential to countering information threats. Resilience-building underpins all the stages of RESIST and is central in all components.

And, of course, successful communications delivery can only happen when teams and individuals have been able to prepare and build their personal resilience. Understanding the complex environment of information threats as a normal routine will enable you to be able to respond with confidence and agility when you are urgently required

Why do we need RESIST?

According to the most recent figures, there are 5.52 billion internet users worldwide, and 5.22 billion social media users. The average person spends almost two and a half hours a day online.[1] Simultaneously, levels of trust in traditional media are falling globally, as is trust in governments. In the UK, only around 30% of people say that they trust the government.[2]

Your job as a communicator is to understand your audiences and earn their attention and trust. This means developing an understanding of the information environment and any barriers that stand between you and your audience.

[1] https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/
[2] https://www.statista.com/statistics/683336/media-trust-worldwide/; https://www.oecd.org/en/topics/trust-in-government.html

Applying RESIST to your work will help you to:

- Recognise information threats, in order to increase resilience to unexpected, coordinated and harmful communication techniques.
- Monitor and handle data processes, in order to strengthen resilience to unanticipated events and support earlier, better responses.
- Share insights to create better situational awareness and consensus around information threats, their risks and likely impact.
- Compile risk and vulnerability assessments that support better decisions on prioritising resources.
- Follow a process for selecting proactive and reactive communications responses that strengthen resilience.
- Utilise evaluation processes that reinforce systems by ensuring learning from previous experiences is captured and used to make improvements.

It is crucial to acknowledge the public's expectation that government responses to information threats are both effective and respectful of freedom of expression. A thoughtful and considered approach, with clear justifications for intervention, is essential to maintaining public trust. Communications should help establish facts to support robust political debate, and never censor or undermine genuinely held opinions and beliefs.

Communication functions should work closely with policy and other colleagues who hold relationships with social media platforms or who assess the tactics, techniques and procedures (TTPs) of threat actors within the information environment to ensure alignment and coherence.

Understanding manipulated, false, and misleading information

There is a category of manipulated, false, and misleading information that is protected by freedom of speech. This is called mis/dis/mal, or MDM for short, as it refers to three slightly different – but closely related – issues: misinformation, disinformation, and malinformation. The important thing to remember about MDM is that, if it is spread by real people during authentic discussions, it is unlikely to be illegal. Lying, exaggerating, creating rumours, and twisting facts are all governed by freedom of speech, except in very specific cases.

The definitions are as follows:

- **Misinformation is verifiably false information that is shared without an intent to mislead:** Sometimes people accidentally share manipulated, false or misleading information. Perhaps they didn't read it properly, or they misunderstood or misremembered what they had read, or they were given the wrong information to begin with. The effects of misinformation can still be harmful.

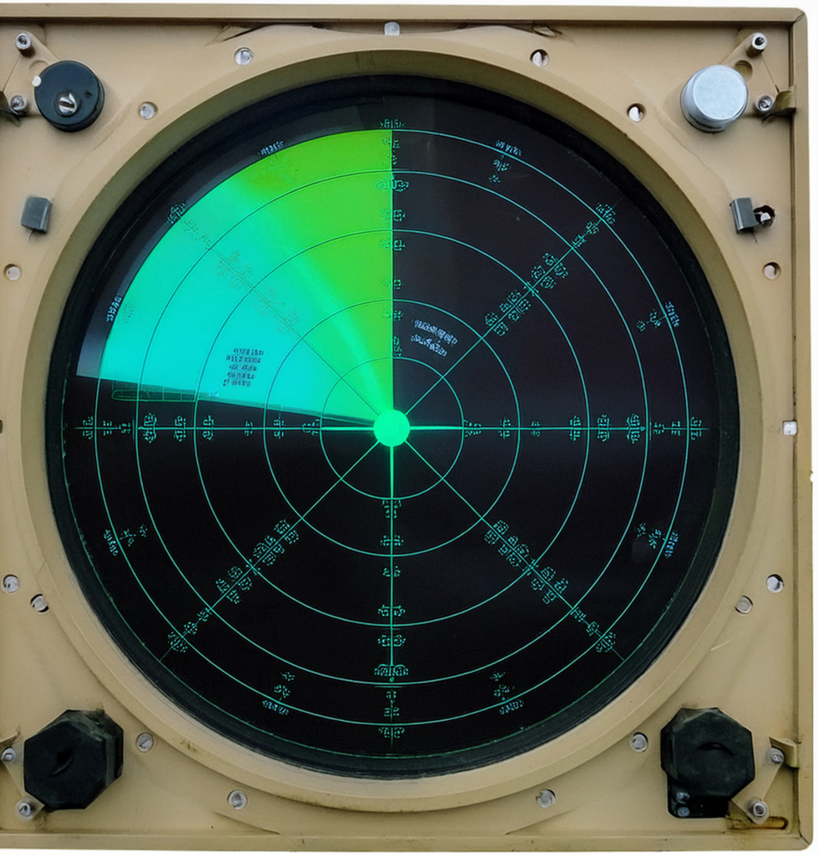
- **Disinformation is verifiably false information that is shared with an intent to deceive and mislead:** People also deliberately spread false or manipulated information. Usually, it is because the individuals and organisations that create it have something to gain from deception. It can often have harmful consequences.

- **Malinformation deliberately misleads by twisting the meaning of truthful information:** Sometimes true or partially true information is used in such a way that it has similar effects to disinformation. For example, facts such as statistics can be misrepresented or taken out of context to support false interpretations.

While MDM is often associated with online content, it's important to recognise that it can also appear in offline forms, such as printed materials. In some cases, online and offline MDM can be used in conjunction to amplify a message.

A democratic response to MDM is to provide access to facts. It is not to censor, ridicule, or blame. It is important for communicators to understand what manipulated, false, and misleading information circulates, and to consider the best ways to reach sceptical audiences with reliable information.

Understanding information threats



While most MDM is legal, things like threats, abuse and harassment, hate speech, and terrorist and extremist content can be illegal. The creation of coordinated, inauthentic networks to spread MDM, while not illegal, can be an indicator of an organised effort to manipulate public opinion. These activities are termed **information threats,** on the basis that they no longer represent the speech of individuals. Rather, they represent deliberate and often sophisticated efforts to manipulate, harm, or coerce others in the information environment.

Sometimes, foreign actors are involved in spreading manipulated, false, or misleading information in the hope that members of the public will believe it and this can lead to further sharing of the disinformation. The foreign hostile state or foreign non-state actor uses deceptive, clandestine, and coordinated methods to undermine national security. This is referred to as **Foreign Information Manipulation and Interference (FIMI).**

FIMI may be defined as a coordinated, deliberate effort by foreign state or foreign non-state actors to manipulate and disrupt a target country's political processes and public opinion through largely non-illegal but deceptive and coercive means, often as part of a broader hybrid strategy.

It will not always be possible to know with certainty whether some identified MDM is part of normal debate or an information threat. The key for government communicators is to assess different indicators and to know when to escalate the issue onwards to units with the expertise to investigate. That means being able to interpret risk. RESIST provides a framework to reach that decision through explaining the different strategic communication stages in responding to an information threat.

Summary

RESIST provides a consistent and effective approach to identifying and tackling a range of manipulated, false, and misleading information that communicators may experience. It provides the building blocks for developing a capability, but not all of the details for implementation. It is best viewed as a conceptual framework that can be used in all types of organisations and is readily adaptable to different contexts and requirements. Each module can be used as a freestanding tool or positioned within a broader process.

Communication departments play a central role in recognising and responding to MDM and information threats. **You will often be the first in your organisation to see these.**

The approach set out in this framework has been proven to contribute to a robust system for recognising and responding to threats and emerging trends in the information environment. RESIST helps you develop routines to make informed assessments of risk, and to share your insights and work with other parts of government. It helps you to formulate recommendations, choose communication responses, and to evaluate their impact. Where necessary, it will also guide you to additional UK Government products.

While the distinctions between domestic misinformation and information threats such as FIMI are useful from a theoretical perspective, in practice they can be hard to distinguish. In many instances, it won't be obvious whether a questionable social media post is entirely false or whether there is malign intent behind it.

It also won't be clear if it is an isolated incident, or indicative of sustained malicious intent. Nor will it be immediately apparent whether the source is an authentic participant in public debate or, for example, a hostile foreign state. This section will introduce you to the important things to look for so that you know how to recognise MDM and information threats.

In this section, you will learn:

- How to identify the components of misleading or manipulated messages.
- Some of the ways that messages fit within and support high-risk narratives.
- How to better understand the behaviour of those who spread problematic messages and narratives.
- How to weigh up the severity of identified MDM and/or information threats.

Recognise

Recognise the narratives

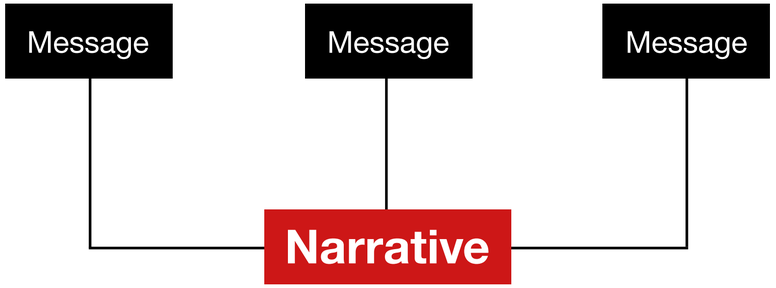
A message is a form of communication aimed at a group of recipients. It could take the form of a social media post, meme, comment, letter, flyer, poster, or slogan.

Is the message an opinion? Opinions are usually subjective, which means that they cannot be verifiably false. Likewise, political discourse is protected. If the message is simply a statement of opinion, you should not treat it as MDM. However, if the opinion is based on verifiably false, deceptive, or manipulated information, it may be worth investigating further.

Messages are the building blocks of narratives. To understand how messages manipulate, you can look to the FIRST indicators (see in **Annex 1**).

Narratives are a form of storytelling that help to explain and shape perceptions of an issue. They are stories designed to influence a target audience. If you see lots of messages on a topic, it is likely that you will be able to identify one or more of the narratives they fit into or help to construct.

Narratives are generally made up of seemingly disparate messages or statements, brought together to tell a particular story. These stories are then more relatable to a broader audience, and can unify groups with different beliefs and interests, making them convenient vehicles for spreading misleading or deceptive ideas. As a communicator, you should familiarise yourself with the misleading narratives that affect your principal areas of work and responsibility.



Narratives give shortcuts to understanding complex issues. They often express things about identity, community, and purpose. They are often not literally true, but rather they carry the aggregated, distilled beliefs of a community built up over time, by many people, across many statements. If you identify a series of manipulated, false, or misleading messages that fit within, or help to construct a larger narrative, that can be an indicator of MDM and/or information threats.

Types of narratives that you should be able to recognise include:

- **Polarisation:** Do the messages explicitly seek to build polarising narratives where there cannot be a middle ground? For example, use of strong emotions that attempt to drown out facts and rational arguments in support of one side of an issue.
- **Social proof:** Do the messages give 'proof' of a larger narrative based on misleading examples? For example, a local incident is used in a manipulated, false, or misleading way as an example within a narrative about an ideological standpoint.
- **Grievances:** Do the manipulated, false, or misleading messages support narratives about genuine grievances that a community might have? For example, the MDM seeks to rile a community that has previously gone through a traumatic experience.
- **Conspiracies:** Do the messages build towards a pre-existing conspiracy narrative? For example, connecting a new event to existing conspiracies about a community or individual.

Remember there can be very good reasons why people believe the things they do. Many beliefs can be fundamental to a person or a community, which makes it more difficult to intervene if those beliefs end up being exploited by threat actors or lead to harmful behaviours. The Government Communications Wall of Beliefs toolkit [3] offers a behavioural science perspective on how to counter false beliefs. In all cases, it can be helpful to take an empathetic view of audiences and to try to understand how and why their beliefs inform the narratives they are susceptible to.

In terms of how deeply ingrained a belief is, it can be useful to consider the following:

- **Beliefs derived from fashions:** "*I believe that electric vehicles are not popular in my area.*"
- **Beliefs derived from recent news:** "*I believe that electric vehicles are expensive.*"
- **Beliefs derived from trusted sources:** "*I believe that climate change is occurring.*"
- **Beliefs derived from norms:** "*I believe that recycling is a good thing to do.*"
- **Foundational beliefs:** "*I am proud to be from my country.*"
- **Sacred values:** "*I believe in a higher power.*"

[3] https://gcs.civilservice.gov.uk/publications/the-wall-of-beliefs/

Recognise the behaviour

On social media it can sometimes be hard to understand whether an account represents a real person, an advocacy movement, a business, or a troll trying to spread certain messages. An account may make claims to be a certain type of person or organisation, or it may give very little information about itself. Since it may not be possible for you to accurately attribute a social media account or set of posts to its owner, it is better to focus on how the account is behaving and how it fits with other accounts in its immediate network.

If the account sharing misleading content is willing to discuss, delete, or correct false statements, it is a strong indicator that there is no intent to mislead. If, however, there are signs that the account is deliberately spreading false content, trolling, or attacking individuals and organisations, the risk is higher that t here may be an intent to cause harm.

If you see lots of accounts posting identical things, this could also be an indication of coordination.

The key question is, does this look like the authentic actions of an individual expressing their beliefs, or does the behaviour suggest inauthenticity?

Some indicators from how the account behaves could include:

- **Suspicious accounts:** Parts of the account profile don't seem to match or appear credible, such as their name, image, or bio, when compared with the account's behaviour.
- **Automation:** The account posts lots of identical content, or makes posts identical to other accounts in the network.
- **Timing:** The release of identical content simultaneously.

Some indicators from the content posted could include:

- **Trolling:** The account exhibits behaviours that may suggest it isn't acting in good faith, for example, engagement is often twisted into opportunities to troll. However, it is important to recognise that robust debate can sometimes appear aggressive without necessarily being insincere.

- **Crowding out:** The account seeks to intimidate, diminish, or ridicule people it does not agree with, for example, discussions often seek to make the situation
so toxic that real people do not want to participate.

- **Targeting vulnerable communities:** The account seems to target a community that has grievances, or is vulnerable to certain types of messaging, for example, the effort seems deliberately provocative, exploits vulnerabilities, or seeks to add fuel to a fire.

- **Doxing** (sometimes spelled "doxxing", means publishing someone's private personal information online without their consent, usually with malicious intent): The account is quick to reveal others' personal information; for example, during a discussion the account makes references to where somebody lives or works.

It is critical to avoid drawing conclusions based on isolated incidents. Many of the behaviours listed above can occur in legitimate discourse. It is the accumulation of these indicators, along with the broader context and potential for harm, that should prompt further investigation.

The advancement of cheap and accessible AI tools adds an additional set of behaviours that may be difficult to track. Text, images, videos, and audio can all be generated or manipulated by AI. Use of AI does not in itself indicate an information threat but threat actors can augment their existing messaging, narratives, and behaviours with more advanced production and distribution capabilities.

This could mean:

- **Content produced and distributed at greater scale:** You may see more content that is a variation on MDM themes.

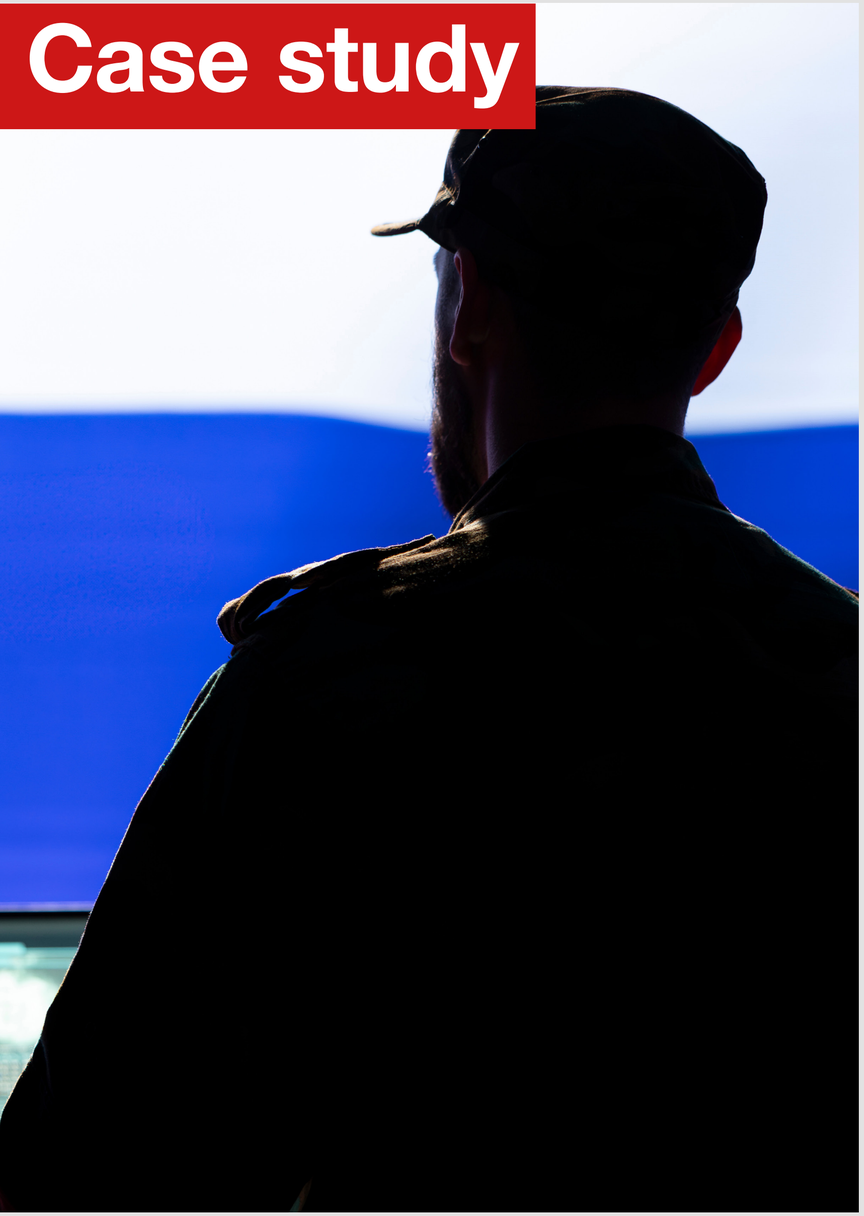
- **Content readily adaptable to more media formats:** You may see the same ideas redeployed across text, images, videos, and audio formats in a coordinated manner.

- **Content easily seeded to more platforms, assisted with the use of AI:** You may notice similar trends in different places, to build the impression of volume. The use of AI to automate workflow processes is creating more robust distribution methods.

- **Content translated to more languages:** You might notice identical or similar content targeting audiences in different countries.

- **Content tailored to narrow audiences:** You may spot examples of content being localised to very niche audiences, which may have been achieved with advanced analysis using AI.

Case study



‘CopyCop’: Fake news to sow geopolitical divisions

A number of Russian-led disinformation campaigns dating back to 2022 have used AI and generative AI to generate and amplify disinformation. Operations including Doppelgänger and False Façade (also known as CopyCop) have used inauthentic media outlets to publish political content relating to France, Israel, UK, Ukraine, and the US.[4] The campaigns have primarily sought to sow geopolitical divisions.

Clone sites of media outlets were created using Large Language Models (LLMs) to plagiarise, translate and edit content. AI-generated images and videos were also used to make sites more convincing.

Operation False Façade comprised a network of at least 23 inauthentic websites that targeted audiences in France, the UK, and the US. Most sites were established within a relatively short time frame, indicating a coordinated launch.

Between May 2023 and July 2024, Operation Doppelgänger published a total of 7,983 articles that reached audiences in France, Germany, Israel, Italy, Latvia, Poland, the UK, Ukraine, and the US.

Content was tailored for EU audiences and relates to divisive issues, including the Russia-Ukraine war and Israel-Hamas conflict. Content on the 2024 US election focused on support for Republican candidates. Bot accounts were used to amplify disinformation.

[4] https://www.disinfo.eu/outreach/our-webinars/11-july-false-facade-and-copycop-two-names-for-a-new-russian-influence-operation/

Recognise the severity

The final step in this stage involves understanding how the messages, narratives, and behaviours fit together and create an impact. An impact can be to degrade the quality of debate, threaten public safety, or to further undermine national interests.

To evaluate how the various indicators you have collected fit together, consider using a matrix:

- **Who is the actor?** What have you been able to determine about the account's behaviour and place within a network?
- **What are their goals?** Are they good faith participants in a discussion or do they seem to have a different agenda?
- **What are their actions?** Are they using their freedom of speech to exaggerate or lie, or are they crossing a line into more harmful activities?
- **How are they doing it?** Are they simply sharing opinions or do you see examples of coordination and more advanced manipulative capabilities?
- **What are the effects of their actions?** Is it poor quality debate, efforts to undermine a reputation, or a potential threat to public safety?

Determining the identity and motivations of those behind MDM and information threats can be complex and time-consuming. While immediate identification may not always be possible during an incident, post-incident analysis can be valuable for understanding the full scope of activity and identifying responsible parties. All analysis must be conducted in compliance with relevant legal frameworks.

Case study

Moldova: Recognising gender and identity-based disinformation

Disinformation campaigns use gender roles, gender equality and sexual orientation to discredit, intimidate, and/or silence women; dissuade them from being politically active and taking part in democratic processes; and undermine gender equality and social cohesion. This often displays an intersectional approach, which deploys and combines further identity-based characteristics such as race, class, and disability to exacerbate and compound harm.

The President of the Republic of Moldova, Maia Sandu, has been the target of extensive gender and identity-based disinformation. A 2024 review of Moldovan Facebook posts containing gender-discriminatory language found that 90% of the posts were targeted at Sandu. Gender and identity-based narratives in Moldova have attempted to undermine public trust in women in leadership positions, while stoking fears around national identity and national security. In this way, disinformation campaigns in Moldova have been found to attempt to undermine Moldova's accession into the EU and weaken Moldova's ties with the West.[5]

Gendered disinformation campaigns often use technology to create fake content, target their victims, and/or amplify their attacks, with real-world impact. This could include posting fake sexualised information, or doctored images and videos that violate what is considered socially acceptable behaviour, or making threats of violence.

[5] https://she-persisted.org/wp-content/uploads/2024/06/ShePersisted-Moldova-Report-ENG.pdf

Summary

Recognising and countering information threats is a critical component of resilience in government communications. This involves understanding the differences between misinformation (false but unintentional), disinformation (deliberate deception), malinformation (true but misleadingly-framed information), as well as understanding FIMI.

By identifying false, misleading, or manipulated messages, communicators can assess how narratives are constructed and spread, including when they leverage AI-driven tactics. Suspicious behaviour by accounts disseminating disinformation, such as coordinated inauthentic activity, or the use of deceptive AI-generated content, can signal broader manipulation efforts.

To prioritise responses effectively, governments can use structured analysis tools, such as a risk matrix, to evaluate the severity of an information threat. These capabilities contribute to resilience by strengthening institutional awareness, enhancing public trust in credible information, and ensuring societies can withstand and counter harmful information campaigns before they undermine democratic processes and social cohesion.

This section will help you to understand:

- How to focus digital monitoring on risks.
- How to use digital monitoring to protect priority issues, tasks, and audiences.

Early warning

Monitoring is about understanding trends

Monitoring of traditional and digital media is a fast-evolving field. It has improved greatly since the days of cutting out interesting stories from newspapers with scissors and storing them in a binder. There are multiple commercial media monitoring services available. Within organisations this can be managed by specialised teams, purchased from analytics companies, or gathered by communicators. They range from very simple metrics (number of likes, shares and comments) to advanced (big data, sentiment analysis and network analysis).

Off-the-shelf tools can provide support to monitoring processes, including:

- **Verification:** Tools that can identify AI-generated or altered text, images, video, and audio.
- **Tracking:** Tools that use AI to identify problematic narratives and track their development.
- **Flagging:** Tools that automatically flag suspicious and/or harmful content.
- **Guidance:** Agentic AI tools (such as chatbots) provide an interface between advanced systems and their users.

AI tools can also support more advanced analysis of content and behaviour. They can help you to understand: the structures and relationships within networks; the methods by which content is amplified (including bots); the origins of content; sentiment; and cross-platform spread of content. In the coming years, these tools will likely become increasingly integrated into normal monitoring workflows, though it is important to note their limitations.

While you likely will already have regular access to some monitoring products, many will not be focused specifically on the risks posed by MDM and information threats.

The existing monitoring you receive should give a good sense of your key audiences, influencers, and of the broader debates that relate to your priority policy areas.

At a minimum, you should use this monitoring to gain an understanding of:

- Digital debates taking place in relation to your organisation and its work.
- The main attitudes held by key influencers and audiences.
- How influencers and segmented audiences engage on digital platforms with your organisation and its work.
- Changes in trends over time.

This knowledge enables you to improve your preparedness for handling MDM and information threats. Spotting emerging trends can buy you a little extra time to prepare for a challenging situation. It will not entirely remove the risk, and in some cases effective early warning will not help much. However, it is better to be prepared by establishing a baseline monitoring routine that gives you a good picture of what your audiences are saying about you, your priority issues, and your services.

Case study

UKHSA: Community engagement

The UK Health Security Agency (UKHSA) is a government agency responsible for health security. Its Risk Communications team monitors vaccine related mis- and disinformation on social media and its Insight team works with communities to understand emerging areas of concern which could pose risks to public health. The organisation recognises that some commentary and content is opinion which differs from UK Government health advice, but does not represent mis- or disinformation.

During a measles outbreak in the West Midlands in 2024, misinformation was circulating among British Pakistani and Somali communities most affected by the outbreak about the vaccines being offered. The misinformation in the West Midlands was identified through community engagement supplemented by social listening. Speaking to members of the community helped uncover concerns and highlighted how MDM was being spread.

UKHSA content was rapidly updated and tailored to address and correct misinformation. Videos were produced with clinicians to engage Muslim parents, and shared with community leaders to post on their TikTok accounts and in community WhatsApp groups. The team worked with TikTok to deliver a campaign fronted by local clinical influencers.

What types of risks should you monitor for?

The National Risk Register (NRR) is a central planning document that contains the government's assessment of the most serious risks facing the UK.[6] If your organisation has already planned for contingencies against a risk register, it can be helpful to reconsider known risks and contingency plans in terms of the potential impact of MDM. This should become a key document for communication teams, enhancing their preparedness for information threats. In addition to using the NRR, there may also be issues that you are already aware of, which should be added to the list.

First, define the risk in a concise and precise way:

- **Description of risk:** What is the problem? Why is it important?
- **Associated narrative(s):** Describe the types of narratives that threaten this area of your work.
- **Types of actors:** What types of actors are likely to be involved in spreading the manipulated, false, or misleading information?

[6] https://www.gov.uk/government/publications/national-risk-register-2025

Second, define exactly how this risk could impact upon your work:

- **Climate of debate:** The risk can negatively affect societal debates about the issue, for example, through polarisation.
- **Reputation:** The risk can negatively affect the reputation of your organisation.
- **Policy areas and goals:** The risk can negatively affect the ability of your organisation to deliver upon its policy objectives.
- **Ability to deliver services:** The risk can negatively affect the ability of your organisation to deliver its critical services. This includes risk to staff safety.
- **Audiences and stakeholders:** The risk can negatively impact upon relationships with these groups, including vulnerable groups.
- **Public safety, public health, and national security:** The risk can negatively affect these areas.

These examples are not exhaustive. The important point is that you begin assessing the risks that MDM can have on your organisation, and prepare for all kinds of information threats. You can also work with colleagues to brainstorm the risks. This can help you to increase your organisation's resilience by, for example, improving campaign planning, better directing monitoring, raising awareness of vulnerabilities, and providing early warning of potential threats.

Remember that the governing rules for data collection can differ depending on your organisation and location. If you are in the UK, consider for example, how Regulation of Investigatory Powers Act (RIPA) and General Data Protection Regulation (GDPR) affect your monitoring. Often, it is more appropriate to monitor hashtags, keywords, and media outlets rather than individuals.

Risk assessment matrix

A matrix can help you systematically evaluate risks posed by different information threats. This template demonstrates how each risk could be consistently assessed, allowing you to plan accordingly.

		Risk 1	Risk 2
Risk	**Description of risk**	Financial destabilisation: The spread of fabricated information about financial institutions' solvency, government economic policies, or market conditions designed to trigger market movements. This could include false claims about bank failures, currency devaluation, or imminent economic collapse.	Community housing misinformation: The spread of false information about local housing policies, social housing allocation, or planning decisions designed to create tensions within specific communities. This could include misleading claims about preferential treatment, or fabricated information about local development plans.
	Associated narrative(s)	• “Major banks are hiding massive losses and are on the verge of collapse” • “Government economic data is being falsified to hide true extent of financial crisis” • “Currency is about to be devalued overnight – withdraw your money immediately” • “International markets are manipulating UK financial systems” • “Digital banking systems have been compromised and savings are at risk”	• “Certain communities are getting priority housing while local families wait years” • “The council is secretly planning to build hundreds of units that will change our neighbourhood” • “Housing benefits are being manipulated to favour specific groups” • “Local residents are being deliberately displaced to make room for others”
	Types of actors	• Hostile foreign states seeking to undermine UK economic stability • Organised criminal networks attempting to profit from volatility	• Local community groups with specific grievances • Local political activists attempting to build support

Impact assessment matrix

This template demonstrates how the impact of each risk can be assessed.

		Risk 1	Risk 2
Impact	**Polarisation**	Medium impact – could create divisions about economic policy	Medium-high impact – could deepen existing community divisions and create new tensions between different groups
	Reputation	High impact – undermines confidence in UK financial system globally	Medium impact – could damage local council or housing authority reputation within specific communities
	Policy implementation	High impact – could force emergency economic interventions	Medium impact – could complicate housing policy delivery and community engagement processes
	Economy	Extremely high impact – could trigger market crashes, runs on the banks, currency instability	Medium-low impact – minimal direct economic consequences beyond potential delays to local developments, and changes in local house prices
	Discrimination	Lower impact – though certain communities might be disproportionately targeted	Medium-high impact – narratives specifically designed to target particular ethnic, social, or economic groups
	Service delivery	Medium impact – could overwhelm banking systems and government services	Medium impact – could affect housing services efficiency and community relationship management
	Vulnerable audiences	High impact – elderly savers, small businesses particularly susceptible	High impact – elderly residents, new migrants, and families in need of housing are particularly susceptible to these narratives
	Public safety	High impact – panic withdrawals could lead to civil unrest	Low-medium impact – potential for localised community tensions but unlikely to cause broader safety issues
	Public health	Low impact – minimal direct health implications	Low impact – minimal direct health implications
	National security	High impact – economic destabilisation threatens national security	Low impact – localised issue with limited broader security implications

Summary

Digital monitoring should be focused on your key priorities. There may be teams in your organisation that provide analysis and insight products. There are also many free and paid tools that can be used to support analysis. You should use combinations of these tools to create a monitoring toolkit that suits your needs.

The purpose of digital monitoring in relation to disinformation is ultimately to help you to reduce vulnerabilities, plan for risk, and protect your priorities. This kind of focused planning can help give you an early warning if information threats appear within your priority policy areas or among key influencers and audiences. The knowledge you develop in these steps should be operationalised in the next step: creation of situational insight.

By the end of this section, you will be familiar with the basic steps required to create and understand an insight report containing relevant information for your organisation.

This section will help you consider:

- The insight you need in the context of information threats and how it can be used to support a timely response.
- How to use the ABCDE framework to generate short-form briefings.

Situational insight

Turning monitoring into insight

Monitoring becomes valuable when it is turned into insight. Insight is a form of analysis that turns interesting data into actionable data. It answers the question, "So what?"

At its core, insight is about understanding audiences to support communication planning.

Insight should be used to:

- Baseline/benchmark over time to show change.
- Identify emerging trends and provide early warning of threats.
- Understand how MDM is distributed to key audiences.
- Generate recommendations.
- Provide support for identifying audiences, as well as developing and targeting messages and campaigns.

Insight is usually presented in reports that are circulated daily, weekly or on an ad hoc basis depending on need. Much of the data can be drawn automatically from monitoring reports or dashboards. A good insight report can be as short as one or two pages: put the most important information at the top and get to the "So what?" quickly.

Your insight product might be the first time that people in your organisation are exposed to digital monitoring data as a basis for analysing information threats. It should be usable as a briefing for special advisers, policy advisers, and senior leaders, so explain things clearly by avoiding jargon and using visuals, where possible.

An MDM insight product should at a minimum include:

- A top-line summary with a short commentary explaining the "So what?" and setting out recommendations for action.
- Sections on priority themes and issues covering:
  - Relevant outputs from your team on priority issues, for example a ministerial announcement.
  - Examples of disinformation relating to these outputs, including where and how it is circulating.
  - Key interactions and engagements, for example is the disinformation being dealt with organically, is it being picked up by journalists and influencers? If so, which ones?
  - Trends and changes in attitudes (and influencers and audiences) over time (this can be combined with any polling data you have).

Collecting and sharing information can be challenging, particularly in multi-stakeholder situations. It is therefore important to be honest about what worked and didn't work in terms of situational insight during recent crises.

A sample insight product is included at **Annex 2**.

Case study

UK Government: Summer riots 2024 crisis response

In the immediate aftermath of multiple fatal stabbings of school-age children in Southport in July 2024, false claims about the assailant spread online. A post on X following the incident was seen by almost four million users. An initial protest drew 150 people but the situation escalated rapidly. Within five days there was coordinated unrest across ten major UK cities.

Social media algorithms and recommendation systems were exploited by networks of far-right influencers. AI-generated images promoted harmful narratives, with users sharing messaging across platforms to maximise reach.

The digital amplification was unprecedented. Rioters used messaging apps to coordinate protests.

UK government departments worked closely together to identify the harmful mis- and disinformation narratives which circulated online. The Department for Science, Innovation and Technology focused on working with the major social media platforms to tackle content contributing to that disorder. This included proactively referring content to platforms which were assessed as likely to violate their terms of service.

Using ABCDE to create briefings

Insight should be actionable and shared within your organisation to policy officers and decision-makers. Recently adopted by NATO,[7] one method of composing briefings that can be coherent to non-specialists and broad audiences is known as ABCDE.[8]

It covers:

- **Actor:** What kinds of actors are involved?
- **Behaviour:** What activities are exhibited?
- **Content:** What kinds of content are being created and distributed?
- **Degree:** How and to what extent is the content being spread?
- **Effect:** What is the overall impact of the case and whom does it affect?

ABCDE can be used to structure longer analyses (for example, a paragraph or a page on each component) but is also useful for short briefings. Anything from a few words up to a sentence for each letter of ABCDE can capture the essence of an information threat for nonspecialists to quickly understand. For example:

*"**ACTOR** (such as, accounts linked to a country) is exhibiting **BEHAVIOUR** (for example, coordinating bots) to spread **CONTENT** (for example, misleading narratives) to the **DEGREE** (for example, currently 500 inauthentic reposts and 1,000 authentic shares) with the **EFFECT** (such as, disinformation that can cause risk to public safety)."*

A sample briefing product is included at **Annex 3**.

[7] https://www.nato.int/cps/en/natohq/official_texts_231905.htm
[8] https://carnegieendowment.org/research/2020/09/the-eus-role-in-fighting-disinformation-crafting-a-disinformation-framework?lang=en

Case study

UK Government: General election 2024 planned response

To manage the risk of foreign state interference or harmful AI-generated MDM during the general election, the Joint Election Security and Preparedness Unit (JESP), a permanent function dedicated to protecting UK elections and referendums, put pre-emptive measures in place to enhance analysis capabilities.

Multi-disciplinary teams across government developed a methodology to collaboratively produce periodic joint reports. Analysis focused on narratives and evidence of coordinated inauthentic behaviour, and the potential risk posed by AI-enabled disinformation or deepfakes targeted at our democratic processes.

Thresholds to determine when a cross-government response would be required were agreed ahead of time.

Whilst the election was not without incident, we did not see any serious information incidents that undermined the integrity of the election. Analysis was effective, and the government remains well prepared to ensure the integrity and security of UK democratic processes.

Summary

The goal of an insight product focused on information threats is to share the early warning signals you have captured from digital media monitoring with the people who need a briefing. As with all monitoring, it can also be used in long-term planning; for example in an annual report or as part of a campaign evaluation. It should be short, clear, and to the point. ABCDE can help to structure short insight briefings to ensure the reader quickly grasps the “So what?”

This section will help you answer the following questions:

- How do I prioritise?
- How do I escalate the most prioritised issues?
- How do I express uncertainty?
- How do I prepare to act?

Impact analysis

Prioritisation

There are some structured analysis techniques that include questions that can help to guide your assessment of the MDM and/or information threats you have identified through monitoring and insight. They can be used to help you decide whether a communication response is required. The following steps will help to unpack decision-making and avoid “gut feeling” reactions.

So far, you have learned to recognise:

- **The building blocks of the messages:** Normally, your first encounter with MDM is in individual messages likely displaying some of the FIRST indicators (see in **Annex 1**).

- **The overarching narratives:** How do the messages contribute to visible narratives? You might see examples of narratives that polarise; offer social or objectively false proof; inflame grievances; or contribute to conspiratorial thinking.

- **Patterns of behaviour within the debate:** Consider whether you see messages and narratives fitting within a pattern of behaviour that seems inauthentic or coordinated, such as coming from suspicious accounts; showing evidence of automation, organised trolling and crowding out of alternative voices; targeting vulnerable users and communities; or doxing.

If you are still concerned by something that you see, make detailed notes about the actor(s) who are spreading the MDM. Use the questions in the severity assessment table below to structure your thinking, and if possible express the likelihood of your assessments. Below is a fictional case study that helps to show what types of notes would be useful.

Severity assessment

You can use what you have found to weigh up the type of information threat you are seeing. In most cases, you won't have a single authoritative answer, so treat what you have found as indicators that help to build a picture of a phenomenon that you can only see part of.

With the severity assessment example above, the judgement would land somewhere between MDM and harmful speech, since there seems to be the potential for more aggravated activity, depending on how the issue develops. This can help to prioritise the issue and to begin to define your organisation's role, responsibility, and communication response.

Who is the actor?	**What can you determine about the account's behaviour and place within a network?** The most central actors seem to be representatives of a local parental association, however I could not find more information about the group online.
What are their goals?	**Are they good faith participants in a discussion or do they seem to have a different agenda?** They often seem to shift the debate from one topic to an entirely different question, and they seem committed to stirring up the public about this issue.
What are their actions?	**Are they using their freedom of speech to exaggerate or lie, or are they crossing a line into more harmful activities?** Some of them seem to get angry and gang up on people who disagree with them.
How are they doing it?	**Are they simply sharing opinions or do you see examples of coordination and more advanced manipulative capabilities?** Several accounts seem to be working together and keep posting links to some low-quality online local news sites as "proof" of their arguments. They have also created a closed social media group for concerned local citizens and encourage likeminded people to join.
What are the effects of their actions?	**Is it poor quality debate, or something more serious?** At the moment it just seems to be a heated debate, but there is an effort to coordinate a larger group of concerned people around this issue. I don't know why.

Escalation

The different categories can help you to determine the likelihood that the issue should be prioritised and/or escalated:

- **MDM:** This is mostly about lawful freedom of expression and should not be considered an information threat unless it is a persistent risk to your ability to conduct business as usual. Persistent MDM should be added to your early warning monitoring, and you should consider some of the proactive strategic communication options referred to in the following section with the aim of improving the overall quality of debate by providing facts.

- **Harmful speech:** Antagonistic speech inciting harassment or violence, especially if it includes elements of clandestine coordination, can be an indicator of a domestic information threat. To be categorised as such, these information threats must contain the genuine risk of causing harm. In this case, other relevant actors within government should be notified and the issue should be escalated to policy level. Your strategic communication efforts in this area, if appropriate, should be coordinated.

- **FIMI:** If there are indications of the involvement of hostile foreign states or other foreign threat actors, the issue should be escalated to policy level and other relevant government actors. Note that FIMI actors often seek to hide their involvement within domestic debate, so these categories are not always easy to disentangle. Your strategic communication efforts in this area, if appropriate, should be coordinated.

The next step is to weigh up what the information threat means to your organisation, and potentially to other parts of government. Some of the information from the previous section can be added here, whereas some will be about how you think the problem impacts your work.

Confidence in your assessments

It may seem counterintuitive, but assessing information threats is not an exact science. Often, we can only see a small part of an emerging picture. Situations are dynamic and ongoing, and there is rarely enough information to be sure of who is behind activities, what their goals are, and what impact their efforts will have. What we see on social media or in insight reports should be treated as indicators of possible trends rather than firm evidence of public opinion. Therefore, we must consider these activities in terms of risk and likelihood.

Careful use of language is crucial to give a nuanced assessment of what appears to be happening. For example, you may have high confidence that a piece of social media content is disinformation, but low confidence in who is ultimately behind it and why they are spreading it. If you are sure, say so. If not, it is helpful to add a small indicator of how sure you are that something is true. It can be enough to place a letter in parentheses at the end of a proposition: low confidence [L], medium confidence [M], or high confidence [H].

High confidence [H]:
The evidence currently available is sufficient to reach a reasonable conclusion.

Medium confidence [M]:
It is possible to reach a reasonable conclusion based on the available evidence, but additional evidence could easily sway that conclusion.

Low confidence [L]:
There is some relevant evidence, but it is taken in isolation or without corroboration.

Here is an example using this method of assessment:

The tone of debate is harsh and is rapidly deteriorating. Some parts of the debate have moved to a closed group [H].

- We have **high confidence** that the closed group has been created and that some participants in the debate are active there, because (i) other participants have provided screenshots and (ii) several accounts refer to these debates.

Preparing to communicate

Once the previous steps are completed, you should be able to assign a priority level to the information threat. Is MDM at risk of causing harm to the public, or are there indications that it will be ignored? Is FIMI likely to become part of a major international crisis? Can conspiracies endanger life, such as misinformation around extreme weather events? Or is it enough simply to monitor developments?

You may need to develop your own criteria for prioritising information threats based on your specific context. The principle is that the goal, impact, and reach should inform how urgently you prioritise the case.

Keep your assessment outcome-focused, taking into account the colleagues you are engaging with and what decisions and actions your assessment may trigger – for example does what you are seeing represent a significant obstacle to achieving your priorities? If not, it should be a lower priority. The role of government is not to respond to every piece of manipulated, false or misleading information. You should not take on the role of arbiter of truth or moderator of public debate. A prioritised response is one in which there is a clear and compelling need to act.

Example of prioritisation thresholds for the UK Government

This is an example of a strategic tool to guide communicators in assessing and prioritising information threats. It helps ensure that interventions are focused and resources are allocated effectively, by indicating when a response is required.

	Description	Actions	Internal audiences	Tools
High	Significant risk to the public and has a high likelihood of attracting media attention. Much of the evidence is high confidence. Requires immediate attention and escalation.	Make senior staff and other parts of government aware of the issue and its priority. Share insight and analysis. Prepare quickly for a cross-government response.	Senior staff Wider government	Share insight Briefings Prioritise short-term communications
Medium	Negative effect on a policy area, departmental reputation, or a large stakeholder group, and is trending online. Evidence indicates a potential for harm if left unchallenged. Requires a response.	Make senior and policy advisers aware of the issue. Share insight and analysis within the department. Investigate the issue and prepare press lines based on known facts.	Senior staff Policy advisers Monitoring and analysis teams	Insight Briefings Press lines Prioritise short and medium-term communications
Low	Potential to affect public perceptions, for example, about a department's work, and has limited circulation. Evidence is of mixed quality. The debate should be routinely followed but intervention is unnecessary/undesirable.	Share insight and analysis within the communications team. Investigate the issue and prepare press lines based on known facts. Conduct a baseline analysis of the debate and track any changes.	Communications team Monitoring and analysis team	Insight Press lines Baseline analysis Prioritise medium and long-term communication

Summary

The assessment of risk and impact in communications should be tackled in a coherent and consistent way, using common tools to make similar assessments. This section gives you suggestions for approaches that can standardise the assessment of risk and impact, leading to a priorities-based approach to developing communication responses that build resilience to information threats.

Strategic communications

Not all MDM has to be responded to. In many circumstances, public opinion will self-correct, and reacting may inadvertently amplify the MDM. Any public response to false or misleading information that your organisation decides to make should deliver the truth well told.

If you decide to act, there are many options – and combinations of options – at your disposal. This section will outline various options and discuss how to deploy them effectively to minimise the impact of manipulated, false and misleading information on your priority issues and audiences.

This section will help you answer the following questions:

- What are the most important communication principles?
- What are proactive communication options?
- What are reactive communication options?
- How can I improve partnerships and capacity for communication?

Follow communication best practice

The Organisation for Economic Cooperation and Development (OECD) has used the collective expertise and learnings of its members and partners to develop good practice to help address information threats.

'Facts not Fakes: Tackling Disinformation, Strengthening Information Integrity' is a report that sets out an analytical framework to guide policy design.[9]

The report looks at three complementary areas:

- Implementing policies to enhance the transparency, accountability and plurality of information sources.
- Fostering societal resilience to disinformation.
- Upgrading governance measures and public institutions to uphold the integrity of the information space.

From a communication perspective, 'Facts not Fakes' draws on the OECD Principles of Good Practice for Public Communication Responses to Mis- and Disinformation.[10]

These were devised to:

- Inform government policies and communication that resonate with citizens' needs and leverage stakeholders as part of a whole-of-society approach.
- Empower communicators through institutionalised approaches that are public-interest driven and evidence-based.
- Mitigate the spread and effects of mis- and disinformation through building capacity for timely, preventive and forward-looking efforts to respond to problematic content.

The document identifies nine common principles underpinning good practices for how governments can engage with partners across citizens, civil society, and the private sector, based on evidence and interventions observed around the world. It is a good reference for anyone working in public communications on information threats.

[9] https://doi.org/10.1787/d909ff7a-en
[10] https://doi.org/10.1787/6d141b44-en

Proactive preparedness and building resilience

A healthy information environment is dependent on people having access to accurate, reliable and timely information, which enables informed decision-making and democratic participation. Dealing with information threats is therefore often based around the principle of empowering individuals to critically evaluate information and increasing their resilience to manipulation.

Proactive communication responses include:

- **Public information and raising awareness:** Evidence-based content that ensures factual information is provided in the information space on priority policy issues or areas identified as higher risk to reduce the space for MDM.
- **Public resilience building:** Improving media literacy to empower individuals to fact check information, and review the source and its validity, to build resilience.
- **Building trust in public communications:** Adopting a transparent, timely, and whole-of-society approach to support citizens to consume evidence-based information.
- **Counter-brand campaigning against threat actors or adversaries:** Implementing a range of communication activities that seek to ensure a reputational cost to actors or adversaries who persistently spread false, misleading and harmful information.

Case study

Government of Montenegro: Media literacy

Social media is widely used in Montenegro. Yet, when considering exposure and resilience to mis- and disinformation, a national survey by Ipsos found that news disseminated through social networks recorded the lowest level of trust at 32%, while 59% of citizens perceived it as unreliable.

The successful UK government ‘SHARE’ campaign – advising citizens to follow a checklist of Source, Headline, Analyse, Retouched, Error – provided the blueprint for a communication response. In 2024, a campaign was launched by the Montenegrin government to build citizen resilience to false and misleading information and promote media literacy.

A localised version of ‘SHARE’ with adapted messaging, acronym and visual identity was created and tested with the Montenegrin target audience. The media literacy campaign, INFO, educated citizens on the steps they could take to protect themselves from online information threats.

The campaign was coordinated across government departments including the Prime Minister's Office, the Ministry of Culture and Media, and the Ministry of Education, Science and Innovation.

Reactive responses to information threats

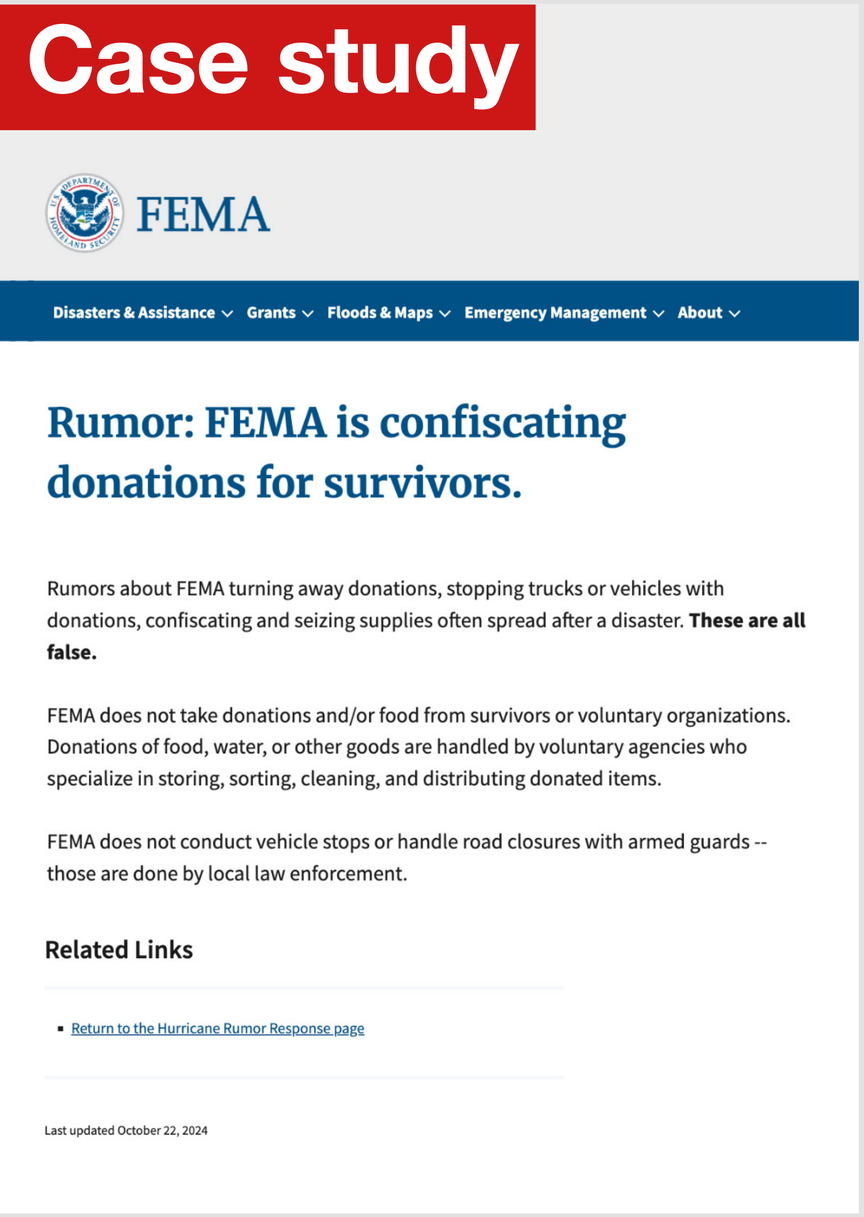
As harmful risks to public safety or democracy are identified, a reactive response must be considered to minimise the impact. In a government context, any response must be considered in close collaboration with policy leads to ensure full alignment. Note that it is also important to consider whether the rebuttal should come from UK Government or from a trusted voice.

Reactive responses include:

- **No communication action; monitoring only:** Taking a strategic decision to not respond to MDM, but conducting real-time analysis.
- **Debunking:** Exposing and countering false or manipulated information by asserting factual information. This carries the risk of amplifying the original false narrative.
- **Counter-narrative:** Promoting a factual narrative without necessarily referring to the harmful narrative.
- **Policy response communication:** Using communications to emphasise policy levers; for example, explaining a decision to levy sanctions.
- **Crisis communications:** Deliver accurate, timely and trusted information to the public about unfolding events.

Case study

FEMA: Hurricane rumour response



In 2024, the southeastern United States was hit by two deadly hurricanes – first Helene and then Milton. As well as the direct impact of the storms, there was also a significant spike in mis- and disinformation online. From false narratives that aid was being diverted to immigrant communities, to conspiracy theories that claimed the storms were the result of the government 'geo-engineering the weather,' disaster relief efforts were undermined.

The Federal Emergency Management Agency (FEMA) is the agency responsible for helping people before, during, and after disasters in the United States.
In response to the unprecedented level of mis- and disinformation circulating, it set up a hurricane rumour response page, with content available in more than 15 languages, which offered advice for individuals to keep themselves, their family and community safe 'by being aware of rumours and scams and sharing official information from trusted sources.'

Capacity-building mechanisms

In addition to proactive and reactive communications, we must also consider the mechanisms for delivering credible communications. These should be built and strengthened over the medium- to long-term. They can be considered as ways to maximise the effectiveness of your communications.

Importantly, many of these mechanisms are dependent upon your relationships with other actors, whether they be from government, within civil society, or part of your audience. In many respects, this is the background work that provides the capacity to co-deliver proactive and responsive communications as and when you need to.

- **Build relationships with trusted voices:**
  Leverage individuals, organisations or entities that are widely recognised for their credibility, expertise and reliability. Trusted voices enhance the impact of communications by fostering trust and confidence in the information being shared. These can be developed into networks and long-term partnerships. Governments often work with one another and with multilateral organisations to amplify communications.

- **Whole-of-society approach:**
  Collaborate with relevant stakeholders, including private sector, media, civil society, and academia. Governments should create an environment conducive to accessing, sharing, and facilitating constructive engagement around information and data.

- **Identify vulnerable audiences:**
  Conduct detailed audience mapping that identifies segments vulnerable to mis- and disinformation. It should be regularly reviewed and used to make channel choices and develop strategies for reaching vulnerable audiences.

- **Identify effective channels:**
  Identify audience reach and penetration by channel. Consider behavioural science approaches (for example, the ‘Wall of Beliefs’ model) for the most effective ways to engage with audiences.

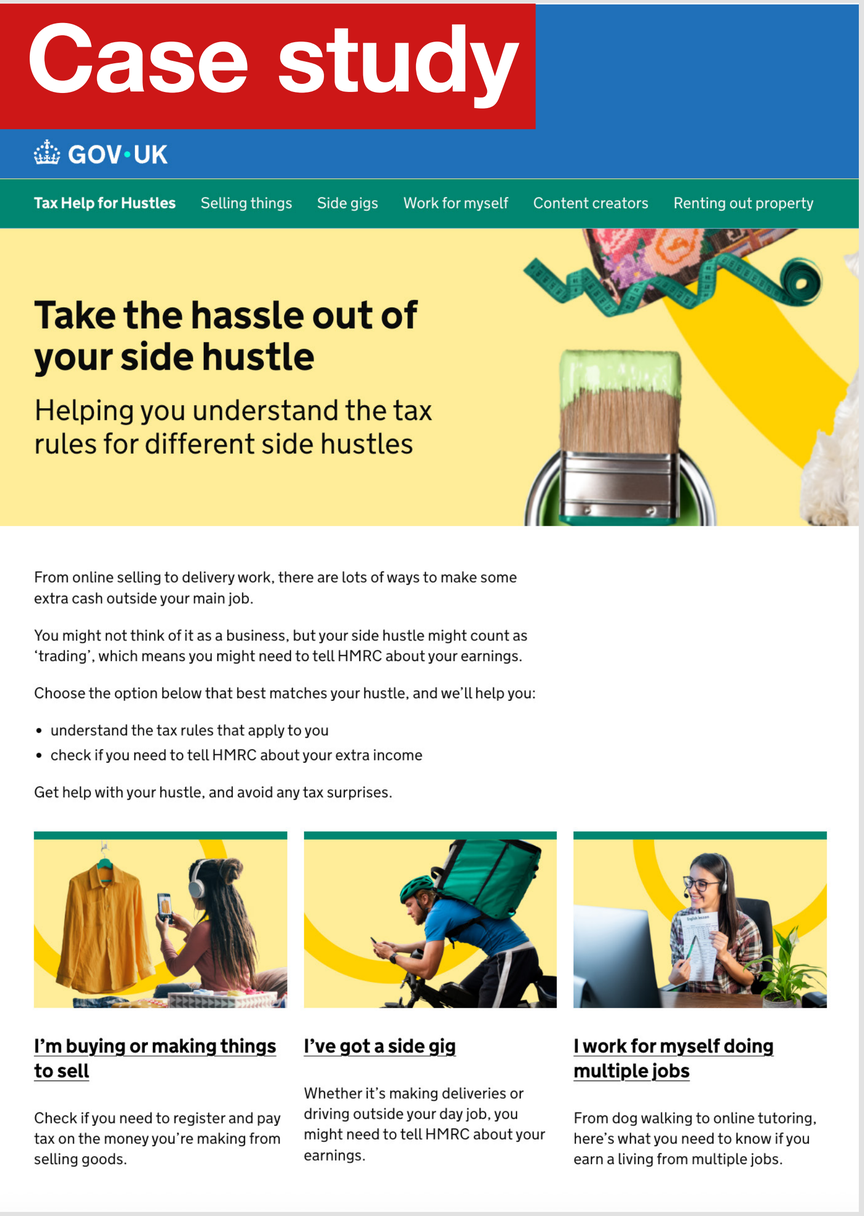
- **Coordination and alignment to policy:**
  Ensure communications are coordinated with policy to determine the most effective response.

- **Build public trust in institutions:**
  Public communications should be transparent, timely, evidence-based and inclusive to build trust in institutions, and reduce the space for information threats to circulate.

- **Adopt a strategic campaign mindset:**
  Develop coordinated, long-term communication strategies using the Government Communications OASIS campaign planning framework to address persistent challenges in the information environment, rather than relying solely on reactive responses.[11] OASIS is a five-step tool to help communicators develop and implement effective campaigns, and consists of Objectives, Audience/Insight, Strategy/Ideas, Implementation, and Scoring/Evaluation. Use of this template ensures interventions are coherent, sustained and aligned with broader objectives.

[11] https://gcs.civilservice.gov.uk/guidance/marketing/delivering-government-campaigns/guide-to-campaign-planning-oasis/

Case study



HMRC: Reactive response to 'Side Hustle' tax MDM

In response to rule changes around digital platforms sharing information with His Majesty's Revenue and Customs (HMRC), misinformation circulated that those selling personal, second-hand items would be liable to pay tax.

The digital communications team noticed a spike in 'side hustle tax' misinformation through their AI-enabled early warning system, built upon NewsWhip's AI digest tool and Brandwatch's Iris AI. The team set up automated alerts for key journalists, newspapers, influencers, as well as volumetric triggers. The system predicted the trend's trajectory and monitored impact on sentiment.

- Moderation of social and local media to proactively rebut misinformation.
- Working with the press team to seek direct corrections to articles.
- Creating a myth-busting video that was shared across channels and amplified by key stakeholders such as eBay as the definitive source of truth.
- Updating GOV.UK through a dedicated 'Tax Help for Hustles' campaign page to provide clear guidance to the public, reducing the risk of misinformation spreading over the long term.[12]
- PR and radio appearances to reach wider audiences with factual information.
- Engaging with influencers on TikTok and Reddit, where HMRC does not have a presence, to correct misinformation.

The department's first response post gained over one million views, and subsequent videos gained over 850,000 views. Martin Lewis, an influential financial journalist in the UK, shared the content, alongside other government departments, and the response was recognised by UK independent fact checking charity, Full Fact. The public started to self-moderate, sharing GOV.UK and verified true content.

[12] https://taxhelpforhustles.campaign.gov.uk

Summary

This section has covered some of the most important principles of public communication and outlined in some detail a range of communication options available to you. Generally, the higher the priority, the more focus should be placed on short-term reactive responses, at least initially. Note that a combination of short-, medium- and long-term approaches may be necessary, depending on the priority of the issue. You can use the Government Communications' OASIS model to plan your campaigns.

In an increasingly complex information environment, understanding the impact of communication activities is crucial. This doesn't just mean measuring success – it's about learning, adapting, and improving communication responses to evolving threats.

By the end of this section, you will be able to:

- Set clear communication objectives in line with policy aims and organisational priorities
- Measure the effectiveness of communication activity
- Systematically evaluate communication response processes

Tracking effectiveness

Why evaluate your communications?

Evaluation serves three broad purposes:

- **Accountability:** Demonstrate the value and impact of your interventions.
- **Learning:** Understand what works, what doesn't, and why.
- **Adaptation:** Make real-time adjustments to improve effectiveness.

However, tracking the effectiveness of communication responses to information threats is difficult. It has to go beyond simple metrics and analytics, and consider real-world outcomes and impacts.

It has two distinct aspects:

- **Evaluating communications.** This section uses the Government Communications Evaluation Cycle to provide a framework for setting objectives, establishing baselines, and measuring immediate outputs and longer-term outcomes.
- **Evaluating the response process.** Even the best communications need effective delivery mechanisms. This section helps you assess and improve your organisation's ability to identify, respond to, and learn from information incidents. It provides practical frameworks for evaluating processes, structures, and capabilities.

By considering both aspects, you can develop a comprehensive understanding of your efforts to build resilience against information threats and make evidence-based improvements to your organisation's approach.

Case study

UK Government: Summer riots 2024 paid media rapid response

Digital communications was at the heart of the communication response to a series of riots which broke out across the UK in summer 2024. To address the increased risk of further breakouts of violent disorder related to the spread of mis- and disinformation, the UK Government ran paid media across Meta, YouTube, TikTok and Snapchat.

Content was targeted to all UK adults and upweighted in areas identified as vulnerable to disorder. The creative combined deterrence messages, with taglines like ‘Violent disorder is not worth it’ and ‘Warning: Actions have consequences,’ with content promoting unity.

The rapid response campaign reached 28.7 million adults via paid media in its first week. One in every eight UK citizens who saw an ad watched it in full (compared to a ratio of one in twelve for typical UK government campaigns). Directive, text-based video messaging was twice as effective as video with sound alone.

Setting communication objectives

The first step is setting clear objectives that align with broader policy aims and organisational priorities. Without clear objectives, responses can become reactive and unfocused.

Objectives should include:

- **Resource allocation:** By enabling better decisions about how to focus limited resources, helping prioritise communication activities and outcomes.
- **Measurement:** By providing the foundation for meaningful evaluation, allowing you to determine whether interventions are actually making a difference.

Common objectives of communication activity relating to information threats include:

- Reaching audiences vulnerable to specific mis/disinformation.
- Directing audiences to legitimate and credible sources.
- Increasing reach and engagement with accurate information.
- Building audience resilience and critical thinking abilities.

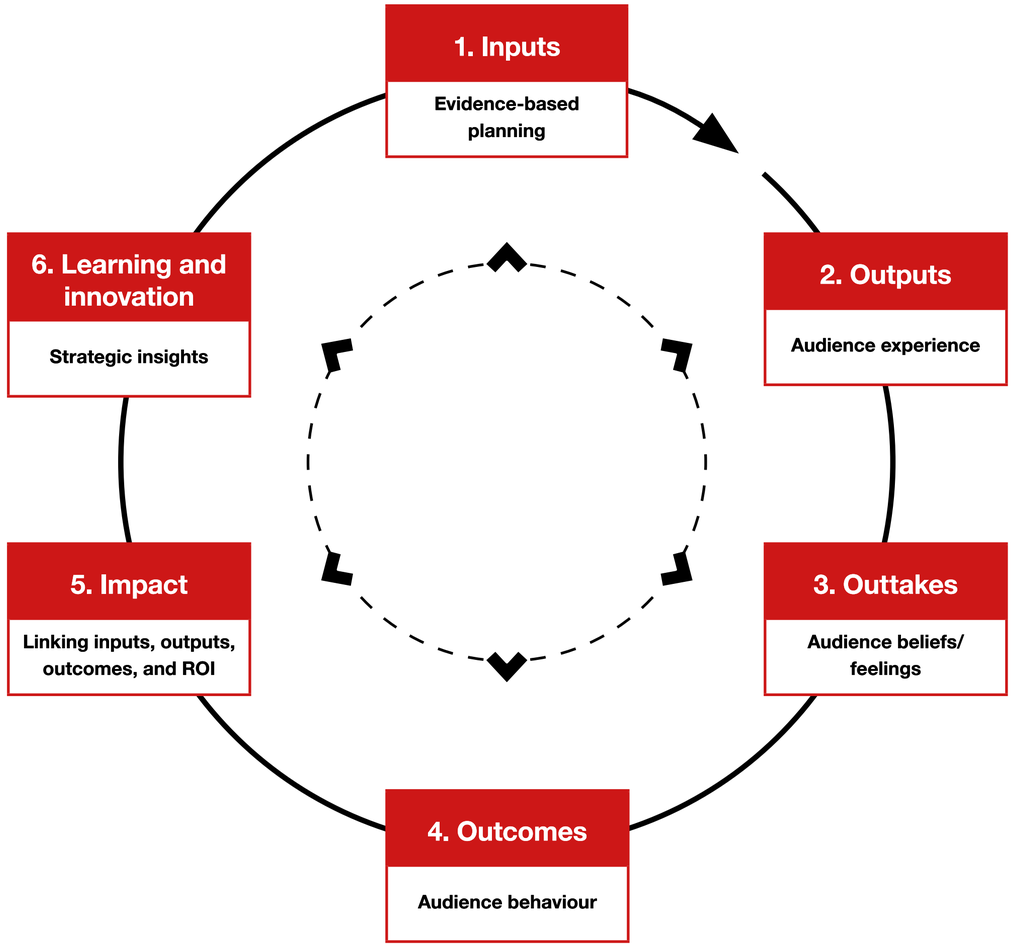
More information on setting objectives for communication activities can be found in the Government Communications OASIS guide to campaign planning.

Once objectives have been set, you need a practical framework for measuring progress at each stage. The Government Communications Evaluation Cycle provides this framework, helping you identify appropriate metrics and establish a systematic approach to measuring change.

The Government Communications Evaluations Cycle

The Government Communications Evaluation Cycle is an agile framework that helps measure the effectiveness of communication activities. Rather than treating evaluation as a one-off exercise, it encourages a continuous process of assessment and learning.

For each stage of the cycle you should establish metrics that will help track progress and demonstrate impact. Your choice of indicators should flow logically from your communication objectives and align with the different stages of the Government Communications Evaluation Cycle.



Using the Government Communications Evaluation Cycle

The table shows examples of metrics that could be used to understand the impact of communications to counter information threats at each stage of the Government Communications Evaluation Cycle.

Stage of evaluation cycle	Example metrics			
1. Inputs	Resources allocated Evidence used to inform strategy Content created Channels selected Partner engagement	**2. Outputs**	Reach of counter-disinformation content Engagement levels with messages Channel performance metrics Volume of content produced Distribution of materials
3. Outtakes	Understanding of messages Awareness of threat posed by MDM Trust in official information sources Use of media literacy techniques	**4. Outcomes**	Information-sharing behaviours Use of fact-checking resources Reporting of suspicious content Changes in information consumption habits Community resilience indicators
5. Impact	Reduced spread of MDM Enhanced public trust Strengthened democratic discourse	**Learning and innovation**	What worked and what didn't New techniques identified Emerging threats spotted Partnership effectiveness Resource optimisation opportunities

Putting the framework into practice

To implement the evaluation framework effectively, you need to establish clear baselines for each metric at the start. You can track progress and measure success against these initial measurements.

Once baselines are established, regular measurement is crucial. This isn't just about collecting data – it's about using these insights to drive improvement in real time. Where metrics show your strategy is not achieving desired results, you can make evidence-based decisions to modify your approach, adjust messaging, or reallocate resources to more effective channels, messages, or targeting.

The iterative nature of this process is particularly important in countering information threats, where narratives can evolve rapidly. Your evaluation should be agile enough to capture both immediate results and longer-term impact, while maintaining focus on your core objectives. When interventions prove successful, these approaches can be documented and replicated. When they fall short, the metrics help identify why and inform improvements.

Remember, while it may not always be possible to draw direct causal links between your communications and observed changes in the information environment, consistent measurement against baselines helps build a clear picture of what works. This evidence-based approach helps communication responses become increasingly effective over time.

Evaluating response processes

Even the best communication strategy needs effective delivery mechanisms.

There are several reasons for systematic response processes:

- **Speed and agility:** Every minute lost in delays or unclear processes gives false narratives more time to spread. Identify and eliminate any bottlenecks that slow down your ability to act.
- **Clarity of roles and responsibilities:** Countering information threats often involves multiple teams, departments, and external partners working under pressure. Without clear roles and accountabilities, this can lead to confusion, duplication of effort, or critical tasks being missed entirely.
- **Resource optimisation:** This work often requires significant resources – both human and technical. Understanding how effectively these resources are being deployed helps you make better decisions about where to invest and how to structure your teams.
- **Partnership effectiveness:** Evaluating response processes helps ensure partnerships – across government, civil society, and the media – are working efficiently and identifies areas where collaboration could be improved. Information threats will often impact or involve multiple teams.
- **Organisational learning:** Each information incident provides valuable lessons about what works and what doesn't. Without proper evaluation, these lessons may be lost, leaving your organisation vulnerable to repeating the same mistakes.

Mapping the response process

The first step in evaluating the effectiveness of your response processes is to map out how your organisation identifies, assesses, and responds to threats.

Begin by identifying all the teams involved in countering information threats. This could include communication teams, analysts, policy experts, and crisis response units. Consider how these teams interact with external partners and subject matter experts. Understanding these relationships is crucial for coordinating effective responses.

Next, document how information flows through your organisation. Create detailed process maps showing how threats are identified, assessed, and addressed. These should include clear decision points, showing who has authority to make different types of decisions and when escalation is required. For instance, a minor incident might be handled at team level, while major disinformation campaigns might require senior political involvement.

Your process mapping should also clearly define who owns each part of the response. This means documenting not just who does what, but who has ultimate responsibility for decisions and outcomes. This ownership structure should align with your organisation's legal and policy frameworks, including any relevant legislation or departmental mandates.

Re-evaluating response processes

Once your response process has been mapped, regular evaluation helps ensure it remains effective and adaptable.

Consider the process you have mapped:

- Does it make sense?
- How consistently is it applied?
- How flexible is it to emerging new threats?
- How quickly are threats responded to?
- How effectively do the stakeholders identified collaborate?

To help answer these questions and generate evidence to inform your answers, consider establishing and tracking indicators. How quickly does your team identify and assess threats? How long does it take to implement responses? These metrics help identify bottlenecks and areas for improvement. However, don't focus solely on speed – the quality of assessment and response is equally important – for instance, the accuracy of threat assessments and the effectiveness of chosen response strategies.

Review how well your teams and directorates work together. Are insights being effectively translated into action? Are the right people being consulted at the right time? Regular process reviews help ensure your process maps reflect reality, not just theory. Your process should be flexible enough to handle variety while maintaining consistent standards.

Finally, maintain a culture of continuous improvement. Regular reviews, after-action assessments, and stakeholder feedback should feed into process updates. Document lessons learned and adjust your procedures accordingly.

Remember, effectiveness isn't just about having a good process on paper – it's about having one that works in practice and that can adapt to changing threats.

Evaluation criteria table

These indicators should be regularly monitored and reviewed against established baselines to track improvements and identify areas needing attention.

Evaluation criteria	Example indicators	
Response speed and efficiency	Time from identification to assessment of disinformation Speed of approval processes Time to implement communication responses	
Decision-making processes	Clarity of roles and responsibilities Consistency in assessments and prioritisation decisions Effectiveness of escalation protocols	
Internal and external collaboration	Information sharing between teams Coordination of responses External partnerships and their effectiveness Senior leader satisfaction with handling	
Resources	Staff and teams Training and specialist expertise Tools Budgets	
Learning	Flexibility in procedures Sharing of best practice Reviews and feedback loops	

Summary

Tracking effectiveness brings us full circle to better understand how to constantly improve strategic communication responses to information threats. It shows why it's important to go beyond metrics and analytics to consider real-world outcomes and impacts.

The Government Communications Evaluation Cycle provides examples of metrics that can be used to track impact. As well as the communication activities, it's also vital to evaluate your organisation's response mechanisms and processes. Robust evaluation processes can reinforce systems and build resilience by ensuring learning from previous experiences is captured and used to make improvements.

Conclusion

Since 2018, RESIST has been used around the world to train, advise, and develop the communication capabilities of governments. RESIST consists of six steps modules that together comprise a full process for building resilience to information threats. It is designed to be adaptive and interoperable: you can choose to use the whole process or just the parts that are most useful. It can be combined with other communication guidelines such as OASIS to guide campaign planning, or used as the basis for training sessions.

This third version reflects recent changes in best practices and acknowledges that the field of countering information threats is ever evolving.

Annex

Annex 1: FIRST indicators

The most common way to first notice MDM is when you encounter messages that draw your attention and raise concerns. Perhaps the tone doesn't sound right, or something about the account makes you suspicious.

At this stage, it makes sense to begin by assessing the content of the message: does it contain manipulated, false, or misleading information? If so, it may be worth trying to understand a little more about how this message fits into the wider discussions.

You should look out for five of the most common components of MDM and information threats. We call these the **FIRST indicators**, because they are almost certainly the first things that will draw your attention. Note that many examples will fit under more than one category.

Fabrication

Is there any manipulated content? For example, a forged document, a manipulated image, or a falsified citation.

- After 7 October 2023, pro-Kremlin outlets alleged that Ukraine had sold Western weapons to Hamas. Fake stories, impersonating credible news organisations, were created and circulated to give this allegation weight. One example was a fake BBC article that claimed evidence had been found of weapons trafficking to Gaza from Ukraine.[13]

Identity

Does anything point to a disguised or misleading source, or false claims about someone else's identity? For example, a fake social media account, claiming that a person or organisation is something they are not, or behaviour that doesn't match the way the account presents itself.

- A 2022 disinformation campaign alleged that Poland was preparing for an invasion of Ukraine, using false claims about identity to lend credibility to the narrative. Fabricated evidence included a purported BBC video and a fake order from the Polish Armed Forces, both designed to mimic trustworthy sources. Photoshopped billboards featuring Jarosław Mika, then general commander of branches of the Armed Forces of Poland, were circulated on Facebook.[14]

[13] https://www.euronews.com/my-europe/2023/10/19/fact-check-has-ukraine-supplied-hamas-militants-with-nato-weapons
[14] https://demagog.org.pl/fake_news/gotovitsya-li-polsha-k-vtorzheniyu-v-ukrainu-dezinformatsiya/

Rhetoric

Is there use of an aggravating tone or false arguments?

- In the aftermath of the 2017 Westminster Bridge terrorist attack in London, a Russia-linked Twitter account amplified an image of a Muslim woman who appeared in one photograph to be looking at her phone as she walked past an injured person on the ground. The post went viral, being retweeted hundreds of times and shared by anti-Islam blogs as supposed evidence of the woman's lack of concern, despite it being one frame from a sequence of many that showed her distress.[15]

Symbolism

- Are data, issues or events exploited to achieve an unrelated communicative goal? For example, historical examples taken out of context, unconnected facts used to justify conspiracy theories, misuse of statistics, or conclusions that are far removed from what data reasonably supports.

- To build support for Russia's illegal annexation of Crimea in 2014, pro-Kremlin media spread narratives that engaged in historical revisionism, including the false claims that Crimea is a natural part of Russia, that Ukrainian control over Crimea was a historical anomaly, and that Ukraine has never been a genuinely sovereign country.

Technology

Do communication techniques exploit technological advantages in order to trick or mislead? For example, coordination between accounts, bots amplifying messages, or machine-generated text, audio and visual content.

- Mis- and disinformation appeared early during the COVID-19 pandemic and spread quickly. This included content related to the origin of the virus, potential treatments or protections, and the severity and prevalence of the virus. Bots on X (then Twitter) spread COVID-19 misinformation, including fake cures and conspiracy theories. Bots criticised the measures imposed to curb the pandemic, expressed disagreement with politicians, or questioned the veracity of the information shared on social media. Search engine optimisation of junk news sites also contributed to the spread of COVID-19 related mis- and disinformation in other online outlets.

[15] https://bridge.georgetown.edu/research/how-russian-bots-instrumentalized-islamophobia-but-dont-just-blame-the-bots/

Annex 2: Sample insight product

Top-line summary:

Following the Health Secretary's announcement, there has been a notable increase in the spread of false and misleading information concerning the Childhood Vaccination Programme. Analysis indicates a network of accounts spreading false claims about vaccine ingredients and concealed side effects, particularly targeting parents on Facebook. These narratives have been picked up by some news outlets which are spreading these inaccurate claims further. The campaign demonstrates sophisticated coordination and is highly likely to impact vaccination uptake rates in vulnerable communities.

Recommendations:

- **Immediate response:** Activate pre-prepared vaccine communication plan to provide factual information about the new vaccine schedule, including ingredients and testing.
- **Clinical engagement:** Brief frontline health workers with counter-information toolkit and Q&A.
- **Media strategy:** Proactive briefings to journalists with scientific experts to address false claims.
- **Platform engagement:** Share coordinated inauthentic behaviour data with major platforms for review against terms of service.

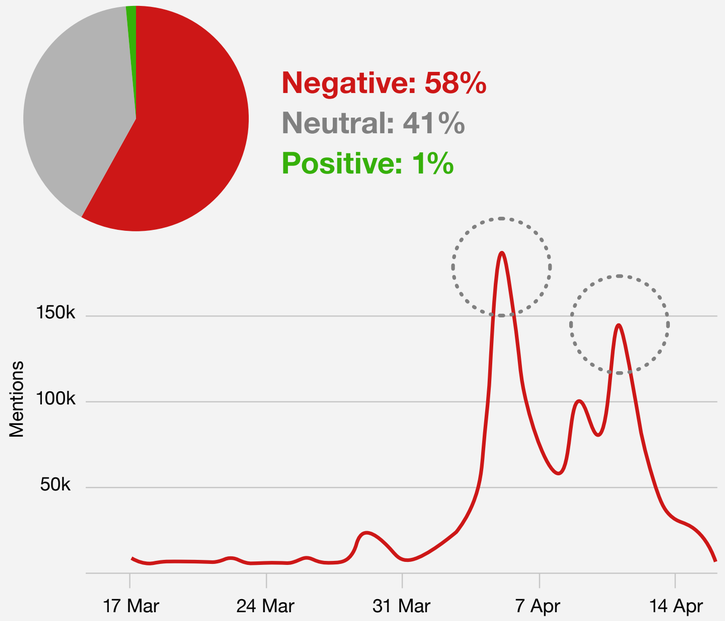
Key MDM narratives identified:

MDM content originated on Telegram and has been amplified on Facebook parenting groups. There is evidence of paid promotion targeting communities with historically lower vaccination rates and coordinated posting, timed to coincide with the announcement on the new vaccine schedule from the Health Secretary.

Narrative	Volume	Assessment	Trend
"New vaccine contains undisclosed ingredients harmful to children"	**1.2k** posts	**FALSE** – Full ingredient list publicly available and verified	**Increasing** (+114% posts in past 24hrs)
"Government hiding data on severe reactions in trial participants"	**0.9k** posts	**FALSE** – All trial data published in peer-reviewed journals	**Increasing** (+52% posts in past 24hrs)
"Vaccine developed too quickly to be properly tested"	**0.2k** posts	**MISLEADING** – Standard testing protocols followed with timeline publicly documented	**Increasing** (+27% posts in past 24hrs)

Volume and sentiment of social media posts:

Social media mentions peaked at 46.81k in the past week (7-13/04), down 28% from previous week. Sentiment has shifted significantly with negative content increasing to 58% (up 16%) and neutral content decreasing to 41% (down 16%), concentrated in parenting groups and communities with historical vaccine concerns.



Key narratives in news and social media:

Mainstream outlets have focused on factual reporting of the schedule implementation, while concerns about vaccine safety are gaining traction across social media channels and some niche news outlets.

- **Announcement of the new Childhood Vaccination Schedule:** Coverage of the Health Secretary's announcement of the updated vaccination schedule, focusing on implementation timeline and public health benefits. (BBC, Reuters, GB News)
- **Parental concerns about vaccine safety:** Outlets and social media users have picked up on the narratives expressing concern about the safety of the vaccines, including false claims about the ingredients and testing of the vaccines. (Facebook)
- **Concerns about NHS capacity:** There has been widespread discussion in response to the new schedule questioning whether the NHS has the resources to deliver the expanded vaccination programme. (Sky, Telegraph)

Official content:

Official posts released to announce the new vaccine schedule, alongside the press conference, have generated 2.3k interactions on social media. The majority of these are positive (68%), however there are a growing number of comments on official posts asking the Health Secretary to respond to concerns about the safety of the vaccines.

MDM content:

- False claims of “undisclosed harmful ingredients” (1.2k posts, up 114% in 24hrs)
- False claims of “concealed trial data” (0.9k posts, up 52% 24hrs)
- Misleading claims about rushed testing (0.2k posts, up 27% in 24hrs)

Network analysis shows 76% of high-engagement posts originated from recently created accounts with coordinated posting patterns.

Annex 3: Sample briefing product

Top-line summary:

Following the Health Secretary's announcement, there has been a notable increase in the spread of false and misleading information concerning the Childhood Vaccination Programme. Analysis indicates a network of accounts spreading false claims about vaccine ingredients and concealed side effects, particularly targeting parents on Facebook. These narratives have been picked up by some news outlets which are spreading these inaccurate claims further. The campaign demonstrates sophisticated coordination and is highly likely to impact vaccination uptake rates in vulnerable communities.

Recommendations:

- **Immediate response:** Activate pre-prepared vaccine communication plan to provide factual information about the new vaccine schedule, including ingredients and testing.
- **Clinical engagement:** Brief frontline health workers with counter-information toolkit and Q&A.
- **Media strategy:** Proactive briefings to journalists with scientific experts to address false claims.
- **Platform engagement:** Share coordinated inauthentic behaviour data with major platforms for review against terms of service.

A	**ACTOR:** Network of coordinated accounts originating from Telegram and targeting parenting Facebook groups, with 76% of high-engagement posts coming from recently created accounts, showing coordinated posting patterns. Evidence suggests targeted paid promotion focusing on communities with historically lower vaccination rates.
B	**BEHAVIOUR:** Coordinated posting activity timed to coincide with the Health Secretary's announcement on the new vaccine schedule. Posts show evidence of strategic amplification and paid promotion targeting vulnerable communities. Network analysis indicates inauthentic activity patterns.
C	**CONTENT:** Content includes false and misleading claims about vaccine safety, such as fabricated concerns about “undisclosed harmful ingredients” (1.2k posts), concealed trial data (0.9k posts), and rushed testing procedures (0.2k posts). This type of content is contributing to increased parental anxiety and has the potential to undermine trust in health authorities.
D	**DEGREE:** High and rapidly increasing reach, with social media mentions peaking at 46.81k in the past week. Individual false narratives showing significant growth rates (up to 114% increase in 24 hours). Sentiment analysis shows 58% negative content (up 16%), concentrated in parenting groups and communities with historical vaccine hesitancy. Some mainstream media outlets have begun amplifying these concerns.
E	**EFFECT:** Risk that this information environment could negatively impact vaccination uptake rates, particularly in vulnerable communities, potentially hindering public health objectives. Potential to erode public trust in health authorities and the new vaccination programme, which could increase the risk of localised disease outbreaks.

Acknowledgements

We would like to thank the many UK Government teams and departments who have contributed to this update to RESIST.

About the author

Dr James Pamment is Director of the Lund University Psychological Defence Research Institute.

About Government Communications

Government Communications is the professional body for people working in communication roles across the UK Government. Our aim is to deliver world-class communications that support Ministers' priorities, improve people's lives and enable the effective operation of our public service.

communications.gov.uk