OCTOBER 2021

A framework for information incidents

Version 1

FULL FACT

About the framework for information incidents

Since 2020, Full Fact has been working with internet companies, civil society and governments to create a new shared model to fight crises of misinformation: the Framework for Information Incidents.

The Framework introduces five levels of severity to build a shared understanding of severe incidents, helping to coordinate timely and proportionate responses to crises. It also proposes a set of the most common challenges that emerge for those trying to find and distribute reliable information and/or tackle bad information, and a set of possible shared aims and complementary responses for organisations to consider when developing a joint response to information incidents.

From March to June 2021 Full Fact ran a consultation seeking feedback on the draft Framework. In particular, the consultation looked at the utility and clarity of the Framework's five level severity scheme and asked for feedback on the set of common challenges, and corresponding aims and responses, and robustness of the methodology. In addition, Full Fact convened a group of UK stakeholders to discuss and improve the use of the Framework in a national context. Full Fact also developed a simulation training exercise based on the Framework, which was delivered to 200 participants at a WHO training conference. This helped to test the practical utility of the Framework with people tackling health misinformation from different industries, and to identify improvements. The acknowledgements contain a list of organisations and independent people involved in the first stage of the project, and those who gave feedback in the consultation period.

We refer to 'misinformation' throughout this document, but this framework is intended to also cover disinformation and malinformation as defined in Claire Wardle and Hossein Derakhshan's 2017 Information Disorder report.[1] In developing this work we have drawn on existing research and analysis listed in the Appendix. We are grateful to the authors of these reports for laying the groundwork to understand these complex issues.

This project was supported in 2020 by a grant from Facebook.

1 Wardle, Claire, and Derakhshan, Hossein, "Information Disorder: Toward an interdisciplinary framework for research and policy making", Council of Europe, 2017

Full Fact
2 Carlton Gardens
London
SW1Y 5AA

 fullfact.org/contact

 @FullFact

 fullfact.org

Published by Full Fact, October 2021

Table of contents

Why do we need a framework for information incidents?

We know that certain events can affect the information environment. That could be by increasing the complexity of accurate information, by creating confusion or revealing information gaps – all of which can result in an increase in the volume of misinformation. This was clearly evident in 2020 during the Covid-19 pandemic, which prompted a slew of intensified measures from internet companies, governments, media, fact checkers, academics and civil society to try and tackle the huge amount of misinformation about the virus.

The subsequent responses showed how fast and innovatively those working to analyse and counter misinformation can respond. But it has also thrown light on the need for greater discussion of the principles behind such measures, of what proportionality means, and on the use of evidence. This will be important for responding to other types of information incidents that may be just round the corner.

This document presents a framework for helping decision-makers understand, respond to and mitigate information crises in proportionate and effective ways. We hope that this framework will enable more collaboration, for example sharing information during and in the run up to incidents, joint planning and evaluation, or increased sharing of capacity and resources.

We have produced this framework with the aim that it is compatible with other analysis, including existing frameworks used by different organisations to spot and guide responses during crises. Analogous frameworks are used in mature industries such as cyber security, or emergency responses from public health bodies and governments.[2]

2 World Health Organisation, "Emergency Response Framework", second edition, who.int, 2017; HM Government, "Emergency Response and Recovery non statutory guidance accompanying the Civil Contingencies Act 2004", gov.uk, 2013

What is an information incident?

An information incident is a cluster or proliferation of inaccurate or misleading claims or narratives, which relates to or affects perceptions of or behaviour towards a certain event or topic happening online or offline. This can occur suddenly, or have a slow onset.

The cyber security industry uses the term 'information incident' primarily used to describe disinformation campaigns, whereas this Framework requires a definition which encompasses accidental or well-intentioned dissemination and sharing of false claims (misinformation), as well as intentional or hostile sharing of false information.

Certain events are likely to trigger information incidents, and to have a substantial and material impact on the people, organisations, and systems that consume, process, share or act on information – toward good, neutral or bad outcomes.

An unexpected incident like a terrorist attack is likely to lead to an increased demand for information and news, but there is often a gap before information is confirmed which may lead to a surge in false information or conspiracy theories. An election might spur polarisation or prompt high profile false claims from figures in authority who are usually trusted by mainstream audiences. In both these scenarios, the baseline information environment shifts: information might be complex, incomplete, or shared or consumed in new ways – both deliberate and accidental.

Based on a comprehensive mapping of recent incidents as well as consultation feedback, we have identified nine categories of events or situations that might trigger information incidents that require responses above and beyond 'business as normal'. These categories are not part of the Framework methodology, but are included here to illustrate situations where one could reasonably expect the information environment to be affected. They are non-exhaustive and are broad by design.

Sometimes multiple events will contribute to the same information incident (e.g. where there is war or conflict it's likely that there will also be human rights abuses). Also, multiple incidents might occur at the same time. It is normal for multiple information incidents to occur relating to long term (perhaps polarising) issues such as climate change.

- **Human rights or freedom of expression abuse.** Information incidents might occur during disrupted peaceful protests; violent public confrontations; long term escalating tension e.g. between regions; mass detainment and/or killings; citizenship or demographic changes.
- **Human-induced violence or conflict** Information incidents often occur in the aftermath of a deliberate attack, and situations with high levels of death or displacement of people.
- **Global or regional conflict** Information incidents may occur during wars or periods of intensified fighting.
- **Ongoing hybrid warfare and disinformation campaigns** may have peaks or moments of intensity that constitute an information incident. Some citizens or audiences may experience on a near-constant basis.
- **Nationally significant political or cultural events.** Information incidents are likely to occur where there is an opportunity to exploit polarisation, such as religious holidays, war memorials or commemorations, national or regional votes. Challenges may vary depending on the country's democratic stability.
- **Infrastructure and economic crises** such as major transport disasters and accidents like explosions; shortages of gas, fuel or food; some hacks and leaks or data dumps; a run on banks, rate fixing and shorts.
- **National/regional health emergencies including pandemics and epidemics.** Also known as an infodemic, a phrase coined by the WHO.[3]
- **Natural or man-made environmental disasters or crises.** Information incidents might occur around extreme weather, manifestations of man-made climate change such as forest fires, and natural events like eruptions and tectonic activity.

3 who.int/health-topics/infodemic

Five levels of severity

Beyond the types of situations where information incidents might occur and prompt use of the Framework, we wanted to understand which incidents were more severe than others, and whether there were common attributes between them.

We have listed indicators to help users establish the severity level of an incident – although not every indicator will be exhibited for each incident. The underlying logic is that different levels of severity should require proportionally different actions, ramped up or lowered as needed.

Measures which are seen as proportionate and reasonable in response to a Level 5 incident should not be the same measures which would be taken in response to a Level 2 incident. By agreeing on a level, different actors can establish a common understanding before considering individual or joint responses.

We have decided on five levels to map increasing escalation of severity.

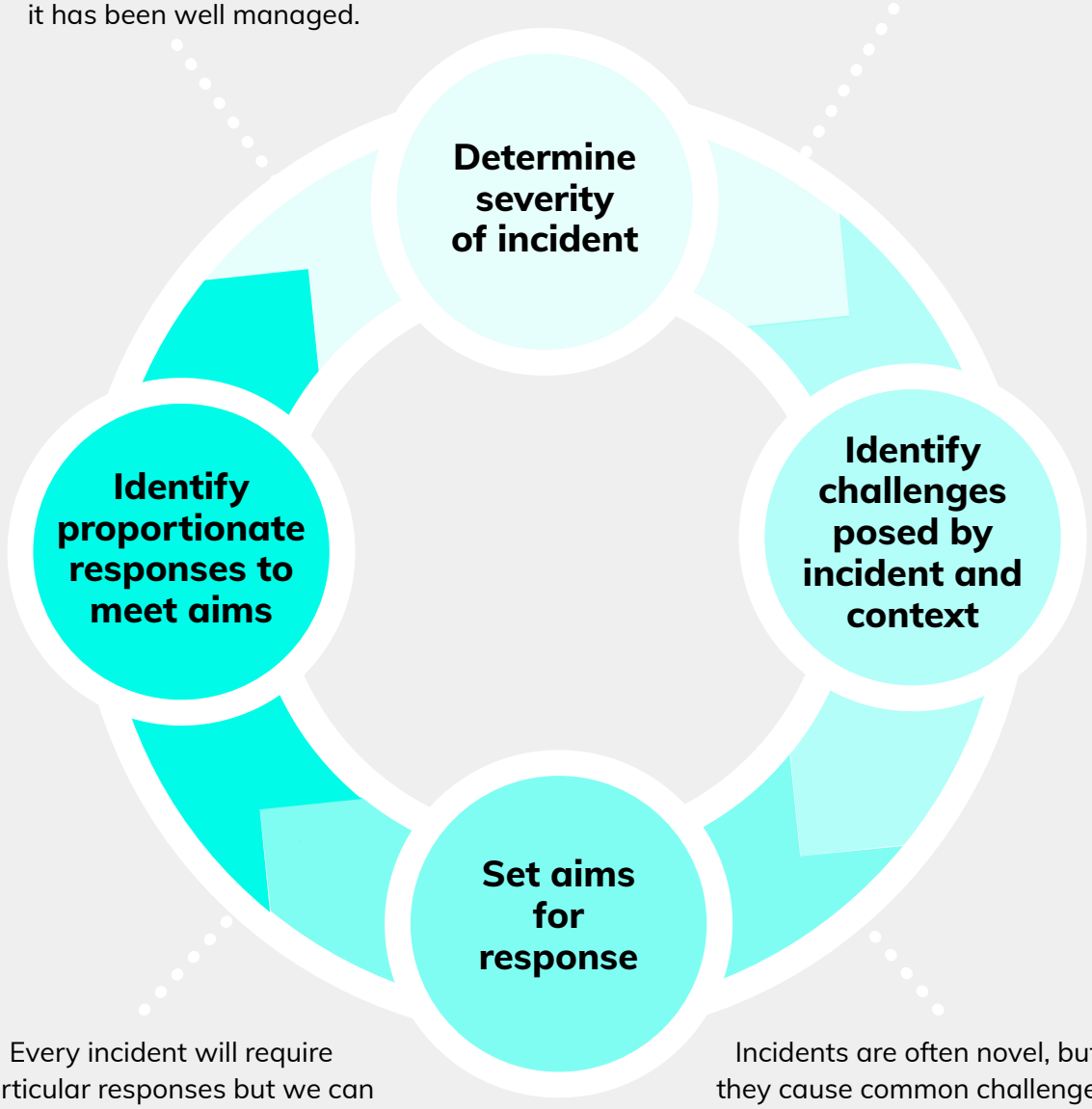
Level 1 is “business as usual” in recognition that some misinformation is, and will remain, a constant fact of life. A world in which there is no misinformation circulating is also likely to be a world with an unreasonable and dangerous amount of surveillance and censorship, and we do not advocate for that scenario. Instead, Level 1 identifies a realistic scenario where there are low levels of misinformation.

Incidents could move between levels over time, whether that is in response to rising severity, for example if it becomes clear that longer-term responses are needed, or where severity is decreasing, such as when an incident is drawing to a close. Early on, it may not always be clear how long an incident will last: this framework aims to offer flexibility to adapt to this reality. Having clear “exit criteria” indicating when an incident can move down in severity will be a priority as we produce the next version of the framework.

Use of the framework in an ongoing cycle

The framework can also be used to show whether an incident is becoming less serious, and to evaluate whether it has been well managed.

A five level framework of determining an incident's severity lets us have shared risk assessments.



Every incident will require particular responses but we can draw on previous experiences.

Different organisations will have different strengths and resources.

Incidents are often novel, but they cause common challenges.

Common challenges means that we can identify common aims in response to many incidents.

Levels of incidents

Level 1

Business as normal, no additional response needed

Engagement: There may be some spikes of engagement and views around certain events, topics or locations but there is low engagement with relevant content overall.

Reach: False claims are not spreading across platforms or languages.

Amplification: Activity around hashtags is very low, and accounts sharing relevant content do not have large followings.

Harm and impact: Content may or may not be potentially harmful.

Effect on resources: Misinformation is circulating at levels considered to be “normal”; there is space to work on long-term goals such as audience resilience and media literacy.

Level 2

Monitor and prepare external facing responses

Engagement: There may be some spikes of engagement and views around certain events, topics or locations but there is low engagement with relevant content overall.

Reach: False claims or narratives are breaking out related to a certain topic or event, on one or more platforms, but not spreading fast/wide or across languages – limited to niche communities

Amplification: Activity around hashtags is very low, and accounts sharing relevant content do not have large followings

Harm and impact: Content has potential to cause harm if the volume of misinformation or engagement grows. e.g. may threaten vulnerable groups or reduce compliance with public safety advice.

Effect on resources: Resources may be diverted to monitor trends but this does not have a significant impact on day-to-day work. There is time to put plans in place to mitigate the growth and effects of misinformation.

Level 3

An incident is occurring, responses begin

Engagement: Content has significantly higher velocity, views or engagement than comparable content would typically have.

Reach: Claim clusters and/or narratives are appearing across multiple platforms.

Amplification: Hashtags/search trends are emerging related to the misinformation; there may be evidence of coordinated inauthentic behaviour with growing traction, and relevant content is being shared by several influential accounts/Pages or by those with higher reach among vulnerable communities.

Harm and impact: Information may be affecting people's decisions or behaviour; Misinformation may potentially damage long-term trust, compliance or democratic participation trends (e.g. undermining health service provision); evidence of minority groups being targeted with hate content or misinformation.

Effect on resources: Day-to-day work temporarily put on hold to tackle sudden proliferation of misinformation, but proliferation expected to tail off quickly.

Level 4

Incident is occurring, coordination/responses ramp up

Engagement: Content has significantly higher velocity, views or engagement than comparable content would typically have.

Reach: Misinformation is present on all major platforms and may be mentioned by mainstream news or discussion shows.

Amplification: Hashtags/search trends are being used to amplify misinformation (intentionally or not); and relevant content is being shared by many influential accounts/Pages or by those with higher reach among vulnerable communities.

Harm and impact: The misinformation might/is resulting in violence or physical harm to individuals or groups and/or significant psychological harm to vulnerable or minority groups; the incident may compound existing conspiracy theories / false narratives. .

Effect on resources: Day-to-day work may be temporarily put on hold in order to implement internal response plans and to enable collaboration with other organisations, including sectors which do not focus on tackling misinformation.

Level 5

Maximum response levels and co-operation required

Engagement: Relevant content has exceptionally higher velocity, views or engagement than comparable content would typically have.

Reach: Misinformation is present on all major platforms and may be mentioned by mainstream news or discussion shows. The incident is global or affecting multiple regions, with the same misinformation often appearing in different languages.

Amplification: Hashtags/search trends are being used to amplify misinformation (intentionally or not); and relevant content is being shared by numerous influential accounts/Pages or by those with higher reach among vulnerable communities.

Harm and impact: The misinformation might/is resulting in violence or physical harm to individuals or groups and/or significant psychological harm to vulnerable or minority groups; the incident may compound existing conspiracy theories / false narratives.

Effect on resources: Response is likely to dominate activity for some time, and collaboration is at a maximum.

Deciding on a level

A few of the reasons to have a shared framework include to improve coordination, establish shared language and promote cohesive approaches. From that perspective, it is beneficial for different organisations to agree together on the level of an incident.

However, it may not be acceptable or appropriate for one or multiple organisations to impose a decision on others. In particular, governments should not declare a level for others. Even in countries without concerns about political dynamics or government being involved in the information flows, a government declaring a level might undermine the action of others and their distinct roles (or perceptions around this). Consultation respondents also felt that technology companies have not yet proved themselves credible to lead a decision-making process like this.

Therefore, a cross-sector group would ideally decide on an incident's severity level, including representatives from civil society such as fact-checkers, local and national government (or former government representatives), press and media, relevant experts and academics and the tech industry. While some organisations or institutions may not wish to commit to something too rigid, flexibility needs to be balanced with operability.

Examples from recent information incidents

Level 2: 5G conspiracies, UK, April 2019

When conspiracy theories about 5G technology in the UK first began to emerge in April 2019, Full Fact highlighted a distinct lack of official guidance that properly addressed public concerns.[4] Claims and narratives suggest that 5G is harmful because signals are more powerful than those that preceded it, e.g. flocks of birds dying, councils cutting down trees that had been harmed, workers wearing hazmat suits to install technology. Level 2 characteristics included:

- False claims and narratives breaking out about safety of 5G on one or more platforms, but not spreading fast/wide or across language – limited to niche communities
- Low activity around hashtags
- Potential for harm in the future (e.g. distrust of government on public health), but no immediate threats to public health and safety
- No significant impact on resources, time to put plans in place if situation escalates

Public health information about the safety of 5G rollout was not improved at the time when Full Fact identified this emerging incident.[5] As a result we saw it increase in severity: 5G conspiracy theories merged with Covid-19 ones in early 2020, attracting celebrity endorsements and leading to the vandalisation of phone masts. At this point we would have classified the 5G conspiracies as a **Level 3** incident, particularly as the conspiracies online translated into offline activities. We saw enhanced collaboration between organisations as the UK government, health bodies and mobile infrastructure companies created new materials on the safety of 5G and the internet companies worked to promote that information on their platforms. Many news outlets also ran explainer pieces debunking conspiracies.

4 Rahman, Grace, "Here's where those 5G and coronavirus conspiracy theories came from", Full Fact, April 2020

5 TPFC report 2019, fullfact.org/media/uploads/tpfc-q1q2-2019.pdf#page=32

Level 3: Notre Dame fire, France, April 2019

The Notre Dame church in Paris catching fire in April 2019 almost immediately prompted false claims that the fire was deliberately started, that the chant "Allahu Akbar" was heard at the church and that a Yellow Vest protester was seen in a tower.[6] Authorities quickly suggested the fire was accidental, relating to a refurbishment. This lack of malicious cause, although accurate, left a vacuum for conspiracy theories and hate narratives aimed at non-Christians, particularly Islamaphobic narratives. **Level 3** indicators included:

- High engagement with and views of content related to the fire, some related content with unusually high velocity
- Minority groups being targeted with hate groups and misinformation.
- Claim clusters and/or narratives are appearing across multiple platforms.
- Day-to-day work temporarily put on hold to tackle sudden proliferation of misinformation, but proliferation expected to tail off quickly.

Organisations moved quickly to share the information that the authorities released about the true cause of the fire, but it took a significant amount of time for that information to permeate given the amount of misinformation online. Some continued to believe the conspiracy theories: in November 2019 a French man set fire to a mosque and shot two men. He told investigators it was an act of revenge for the Notre Dame fire.[7]

Level 4: Afghan refugee crisis, Turkey, August 2021

In August 2021, public debate in Turkey about the ongoing migration flow into the country flared up as Afghans fled the advance of the Taliban. Afghan migration was already a sensitive subject among Turkish society following high immigration of this migrant group to Turkey between 2011-2019. In this context, visual materials circulated on social media during the US and NATO withdrawal contributed to information disorder, with people on social media blaming the West for allowing Turkey to be "invaded" by refugees. **Level 4** indicators at the time included:

- An attempted pogrom against refugees taking place in Ankara on 11 August, fuelled by misinformation and hate speech targeting Syrians.[8]

6 Funke, Daniel and Benkelman, Susan, "5 lessons from fact-checking the Notre Dame fire", Poynter, 2019

7 Abdelaziz, Rowaida and Robins-Early, Nick, "How A Conspiracy Theory About The Notre Dame Cathedral Led To A Mosque Shooting", HuffPost, 2019

8 Euractiv with AFP, "Pogroms attest of growing anti-Syrian sentiments in Turkey", Euractiv, 2021 euractiv.com/section/global-europe/news/pogroms-attest-of-growing-anti-syrian-sentiments-in-turkey

- Velocity of both true and false online content on Afghanistan and Afghan migrants in Turkey was much higher during this period..
- A nationalist politician with 1 million+ social media followers, who previously shared misinformation on Syrian refugees, also shared anti-Afghan content.
- Hashtags which became commonly used included #AfganlarıAlmayın (don't take Afghans), #Afganlar (Afghans), #istila (invasion), #işgal (occupation).
- Day-to-day work was temporarily put on hold in order to plan a specific response to this crisis.

Fact checking organisation Teyit called a staff meeting to tackle the increased demands from its audience and to systematise its editorial prioritisation specifically for fact checks related to the crisis.[9]

Level 5: Start of Covid-19 pandemic, February 2020

This is the only incident to date we would classify as a **Level 5**. As well as global lockdowns and economic crises, the pandemic prompted a slew of measures attempting to grapple with newly challenging types and extremely high volume of life- and health-threatening misinformation. The characteristics that put this incident at the most severe level include:

- Relevant content has exceptionally high velocity, views or engagement than comparable content would typically have.
- Misinformation is present on all major platforms and may be mentioned by mainstream news or discussion shows.
- The misinformation is and may continue to result in physical harm to individuals or groups, and compound existing conspiracy theories.
- Response is likely to dominate the activity of those working to counter misinformation for some time.

The response to this was immediate but was, at the beginning, inconsistent, uncoordinated and required many organisations to create new emergency procedures and work internationally at a scale not seen before. Particularly as it became clear the incident was here for the long term, organisations had to reconsider protocols, funding structures, the deployment of resources and response policies.

9 Thanks to Teyit for providing extra information on this case study.

What common challenges exist across incidents – and how should we respond?

Every incident is unique. But in many cases, common or predictable challenges will emerge for those trying to find and distribute reliable information, or tackle bad information. Many incidents threaten freedom of expression and make it harder for accurate information to be circulated, create an unclear or quickly changing situation, or create lasting problems after the initial incident such as breakdowns in trust.

The existence of these common challenges implies that in some cases it should be possible to plan in advance how to respond and, theoretically, to consider joint aims in advance. It is likely that organisations will prioritise aims differently; it is right that different organisations have different strengths and specialities and this should continue. But we believe that by communicating about intentions (and expectations), finding shared terminology, and even harmonising plans in advance where possible, more effective action can be taken to mitigate the pressures of crises with efficient, credible responses from all actors with the ability to do so.

Sometimes responses may be applied across different timelines: there might be a combination of short term tactical measures and longer term strategic measures.

The prior identification of challenges and aims also points to the necessity of dialogue with other actors that have an interest in a good outcome whether as a directly impacted group or community of practice or specialist organisation.

Challenges across different incidents with aims and possible responses

1 **Threats to freedom of expression**

Challenges: eg when:

- **Lack of independent scrutiny of laws, moderation policies and norms that allows for censorship creep**
- **Unprecedented use of technology to reach large audiences without the ability to independently scrutinise**
- **Lack of protection for minority/vulnerable communities who may find it dangerous/difficult to speak out**
- **Suspected or known foreign interference**
- **Lack of protection for whistleblowers**

Aims: Design responses that are demonstrably proportionate to clearly identified harms, and open to informed debate and discussion

- Provide access to engagement, trends, and advertiser data to enable independent research on the impact of responses
- Evaluate the effectiveness of counter-misinformation efforts and publish learnings
- Provide metrics on takedowns and government interactions with tech platforms
- Enable independent experts to scrutinise AI recommendations
- Consistent use of procedural and audit systems to enable regular transparent reporting

2 **An unclear or quickly changing situation**

Challenges: eg when:

- **Lack of insight into type and scope of misinformation and/or movement of content between platforms**
- **Unhelpful duplication of efforts among organisations**
- **Breaking news which is peddling unconfirmed information: this increases urgency to respond**
- **Contradictory statements from authorities, experts, public health officials**

Aims: Towards a shared assessment of the situation and complimentary responses

- Share monitoring and verification information between trusted experts
- Use of local, on-the-ground sources
- Support smaller platforms to share trends data to help predict when narratives/ claims might move to mainstream platforms
- Brief media and other mainstream sources of information to reduce risk of amplification and stop dissemination of harmful information

3 **Difficulty disseminating or communicating information**

Challenges: eg when:

- Low baseline knowledge of key issues among public, politicians and media
- Low statistical literacy among public and media
- Language barriers between those trying to communicate and those affected by misinformation
- Accurate information is not contextualised or adapted for certain groups
- Topics are complicated or highly technical
- Audiences overwhelmed and find it hard to judge content in the decontextualised format of news feeds
- Intense partisanship/emotive topics make it harder for accurate information to be believed
- Lack of credibility with target audience
- News deserts
- Related disinformation campaigns targeted at multiple audiences simultaneously

Aims: Good information reaches both affected groups and the wider public, and key information is communicated effectively by trusted figures

- Identify and engage with appropriate trusted voices to disseminate information
- Promote relevant impartial or official sources of information
- Disseminate information to pre-empt belief in emerging conspiracy theories
- Make it harder to find harmful content in search
- Expose consumers of high volumes of harmful content to counter messaging

4 **Information vacuums and uncertainty**

Challenges: eg when:

- Information is partial, allowing for distorted reporting and discussion
- New information must be produced, leaving a temporary gap
- Official advice is changing quickly or official sources backtrack
- The future is unknown so unfounded claims of certainty gain traction

Aims: Ensure availablity of reliable information from authoritative sources, and that any limitations are clearly communicated

- Funding and resources for statistical offices and impartial information providers
- Horizon scanning to ensure information is adequate for future public decisions
- Strengthen and support impartial journalism
- Transparently explain why information or advice has changed

5 **Damaging behaviour by influential public figures**

Challenges: eg when:

- **Repeat false claims or make conflicting statements Cast doubt on accurate information**
- **Censorship of news and public debate (of specific topics as well as in general) or intimidation of other politicians/ opinion leaders**
- **Disinformation and propaganda supported by or originating from state/state-backed actors**
- **Deliberately encourage distrust of mainstream media**

Aims: Provide context to help audience make judgements and promote alternative trustworthy sources of information

- Apply warnings, pop-ups and labels
- Promote alternative coverage from trustworthy media and fact checkers
- Give information and caveats about sources of information being presented

6 **Speed or scale pressures to halt spread of false beliefs**

Challenges: eg when:

- **Volume and speed of information increases beyond resources of human teams to monitor and counteract it**
- **Increased consumption of news encourages media to report insignificant stories as major developments and increases likelihood of mistakes being made**
- **Unintended consequences arise from responses including entrenchment of false beliefs**

Aims: Limit bad information, ensure appearance of relevant corrective information, with a clear plan for scaling:

- If appropriate, reduce circulation of harmful false content and/or address persistent offenders in a proportionate and transparent manner
- Design effective corrective content
- Implement additional verification standards before information is disseminated
- Strengthen moderation enforcement policies
- Invest in burst capacity and systems including support for experts and news organisations
- Work with volunteers to feed AI with marked up data for emerging topics or claims
- Combine technology with human judgement to downrank harmful content and sources which repeatedly promote debunked information

7 **Immediate threats to public order and safety**

Challenges: eg when:

- **Public order and safety is dependent on the public understanding information accurately**
- **Disrupting credibility/ability of frontline or aid workers to deliver services**
- **Communication from affected communities and first responders is compromised or ignored**
- **False information creates potential for physical harm through violence or hazard**

Aims: Build audience resilience, and communicate and debunk effectively:

- Adapt or contextualise information to reach target / affected audiences
- Identify and engage with appropriate trusted voices to disseminate information

8 **Lasting longer term impacts of incident/s**

Challenges: eg when:

- **The incident spawns or entrenches conspiracy theories or myths which outlast the incident**
- **False narratives are repeated over years and create hard-to-shift public misperceptions**
- **Radicalisation is evident**
- **Loss of credibility of government/authorities**

Aims: Build audience resilience, and communicate and debunk effectively:

- Cross sector investment in effective communication of information
- Increase audience awareness of and ability to identify bad information
- Research and fund effective teaching methods for information literacy, and evaluate existing information literacy programmes
- Work with schools, universities and qualifications bodies to ensure critical thinking and information literacy curriculums are effective and regularly evaluated and updated
- Targeted education of influencers around certain topics

How do severity levels map across challenges and responses?

It is clear that different levels should require different responses. A proportionate measure to address a Level 4 scenario might not be proportionate in a Level 2 scenario. In a Level 1 situation with normal misinformation flow, the risk is likely to be low in the immediate term.

As we have set out above, **Level 1** is intended to represent the regular, day-to-day environment that organisations work in. In this situation there is no specific incident either currently happening or on the horizon, and therefore no need to do anything beyond business as usual. In this situation organisations can work on long-term priorities such as building media literacy resilience, responding to the topical news of the day, trialling new products or interventions, undertaking research, and planning and preparing for when incidents do occur. Each organisation will have their own priorities and aims to achieve (and will communicate with other actors in common spaces and forums on shared concerns accordingly).

Levels 2-5 represent when something is happening outside of the norm. Our current thinking is for users of the framework to select a challenge (outlined above), and then to identify which of the corresponding aims is likely to mitigate or resolve this challenge. We propose that different severity levels will not affect challenges and aims, but that responses can be calibrated and adapted to ensure they are proportionate to the severity level.

We have outlined possible responses to the challenge of dealing with information vacuums and uncertainty. The suggested responses are not mutually exclusive to each other, for example **Level 4** responses might be put in place in addition to **Level 3** responses.

Challenge: Information vacuums and uncertainty

Aim: Ensure reliable information from authoritative sources is available and that any limitations are communicated.

Level 2 responses: When misinformation is breaking out on smaller platforms which may be polarising, or threaten vulnerable groups:

- Identify topic-relevant information producers and influencers
- Remind public figures to avoid making claims of false certainty
- Establish when information might change quickly and what is needed to keep messaging clear, including how to communicate uncertainty
- Identify potentially problematic gaps in public information and ways they could be filled within existing work plans

Level 3 responses: When misinformation is moving into the mainstream, but the incident is likely to be short term:

- Increase information-sharing with fact checkers and media to enable quick effective rebuttals
- Provide additional relevant information from authoritative sources
- Increase flagging of most viral claims to platforms and authorities
- Communicate to intermediaries where information gaps currently exist
- Avoid creating/promoting speculation about the cause of the incident and work rapidly to put out accurate clear statements
- Transparently explain why information or advice has changed or where there are gaps

Level 4 responses: When misinformation is being amplified, there may be evidence of coordination, and there is potential of physical danger:

- Promote materials that debunk false claims of certainty and link to official sources and accurate news
- Provide access to data on the incident to enable learning and improvement
- Strengthen and support impartial journalism
- Update and introduce policies transparently and collaboratively
- Increase funding and resources for statistical offices and impartial information providers to fill information gaps
- Increase collaboration with organisations to identify false claims and disseminate reliable information
- Coordinate with others to ensure messaging and advice is clearly heard
- Promote best practice for communicating about uncertainty

Level 5 responses: For where there are high volumes of misinformation spreading very fast and likely to cause significant human harm:

- Maximum monitoring and information sharing between all relevant sectors
- Emergency support for impartial journalism and independent research
- Increase funding and resources for statistical offices and impartial information providers to fill information gaps
- Commission real time or rapid research into trends, misperceptions and behaviour which can be applied to learn and improve responses
- Inter- and cross-industry collaboration to create new policies and products to respond to the situation

In the next version of this document, we intend to create a workflow to articulate the different response levels for the other challenges. This would then be converted into an interactive tool that enables users to select challenges and draw up a basic menu of aims and responses – forming a discussion document for planning with team members and other colleagues.

Case studies

Case study 1: Covid-19 pandemic

While Covid-19 was first identified in December 2019, it was not until early February 2020 that much of the world realised that there was a significant incident underway. It took a further few weeks for organisations to realise the extent of the information crisis that was underway. If this framework had been widely operational before the pandemic, the incident could have played out as follows.

Governments around the world begin to implement measures to try and restrict the virus. Recognising that an incident is occurring, representatives from internet companies and government, civil society, health experts and news media meet to discuss. The group recognises that:

- There is clearly a lot of confusion and misreporting, both online and offline, on the symptoms of the virus, how it can be caught and passed on and the new restrictions on socialising and movement.
- Conspiracy theories around the origin of the virus are growing in popularity and translating into anti-Chinese sentiment.
- The incident is global, misinformation is being shared in multiple languages and across borders and engagement is only rising.
- Organisations are struggling to keep on top of how the accurate information is evolving, and tracking the misinformation being shared. It is clear that this cannot be effectively mitigated by individual organisations working alone.

After deliberation and dialogue the group decides this is a **Level 5** incident. This is the first time that **Level 5** has been triggered. The key challenges are identified as:

- Difficulty disseminating or communicating complex scientific information.
- Information vacuums and uncertainty.
- An unclear and quickly changing situation, with contradictory explanations and changing scientific advice.

In addition, the danger of lasting long-term impacts and the pressure of needing to work at speed and at scale are recognised. Academics and civil society point out that there is also a risk of threat to freedom of speech from overzealous new content moderation policies taken in response to these challenges. The number and variety of challenges contribute to the agreement that this is a Level 5 incident. Three aims are chosen as priorities:

Aim 1: Ensure reliable information from authoritative sources is available and that any limitations are communicated

- Government and health bodies have responsibility for collecting and producing accurate and reliable information.
- The internet companies take significant steps to provide information to users, from redirecting search results to providing proactive information boxes at the top of the newsfeed.
- Fact checkers and news organisations seek to communicate the limitations of any information and provide easy to understand explanations of the data available.

Aim 2: Work towards a shared assessment of the situation and complimentary responses

- Researchers, internet companies, governments and fact checkers share information on a regular basis about the situation, including the most common narratives and any emerging claims that may be concerning.

Case study 2: Stem rust crop fail (hypothetical)

At the start of a boiling hot August, reports of stem rust, a serious yield-affecting disease caused by a fungus, begin appearing in Facebook groups for farmers in South East England. Soon, farmers' unions declare an industry emergency: the disease has not appeared in such force since 1955. Theories begin to emerge online: environmentalist groups planted the disease; the government is trying to break the farmers' unions; potatoes and rice will infect your garden.

The summer news lull is quickly replaced by interviews with panic-stricken farmers begging the government to buy up fungicide to protect British crops. Front pages feature close-ups of wheat ears with lurid orange growths next to images of food made from UK grain. As #stemruststockpile starts trending on Twitter, hoaxes soon become reality and supermarket shelves are emptied of flour, bread, yeast and granola. WhatsApp messages urge farmers and their allies to take to the streets and block roads into London.

Recognising that an incident is occurring, representatives from internet companies, government, civil society, farmers' unions, scientists and news media meet and decide the situation is a Level 3 incident It is noted that:

- There is misinformation moving into the mainstream and organic amplification.
- Response plans could be improved through collaboration.
- There is little time to plan; acting quickly is vital to maintaining public order and safety.

The group agrees that the key challenges are:

- Unclear situation: lack of insight into type and scope of misinformation on platforms.
- Limited threats to public order and safety: false information may fuel violence/shortages.
- Information vacuums and uncertainty: unfounded claims of certainty are gaining traction.

The organisations choose several aims to address these problems, with actions for each.

Aim 1: Work towards a shared assessment of the situation and complimentary responses

- Internet companies and monitoring groups agree to a two month period of sharing detailed topic-specific trends data and verification information.
- Fact checkers and impartial researchers begin the first of sixteen coordinated bi-weekly briefings for media detailing the top false claims circulating to the reduce risk of amplification.

Aim 2: Consider targeted measures for affected audiences to see and trust accurate information

- Facebook and Google introduce flash grants to newsrooms and fact checkers to support contextualisation of information to reach affected audiences.
- The government identifies and engages with trusted voices to disseminate information.

Aim 3: Ensure reliable information from authoritative sources is available and that any limitations are communicated

- The UK Statistics Authority diverts resources to ensure that public information on wheat crops and diseases is communicated accurately and corrected quickly.
- The Royal Agricultural Society of England agrees to regular media appearances transparently explaining the background to changes in public advice.
- A consensus is reached that there should be limits on frequency of briefings, media appearances, etc, and an end-date for collaboration, recognising that the incident is of relatively low severity.

Evaluation

One example of evaluation could be that the regulator, The Office of Communications, known as Ofcom, commissions an independent evaluation which looks at organisations' satisfaction with collaboration, tests audience beliefs, and retrospectively analyses patterns in misinformation and online activity. This could be reported back in a debrief meeting two months after the end of the incident.

Appendix

Acknowledgements

Organisations involved in the Framework's initial development

We extend our warmest thanks to those who contributed their time and gracious feedback throughout the first stage of this project in 2020, especially representatives from:

- Africa Check (South Africa/ Nigeria/Kenya/Senegal)
- Boom (India)
- Chequeado (Argentina)
- Department for Digital, Culture, Media and Sport (UK)
- Facebook
- First Draft (UK/US/Australia)
- Google
- International Fact-Checking Network
- Maldita.es (Spain)
- Privy Council Office (Canada)
- Reuters Institute for the Study of Journalism at Oxford University
- Twitter

Organisations who gave feedback during the 2021 consultation

There were five responses from people who do not work for an organisation. Out of these five, three gave details of former employment in the UK National Health Service (NHS), local government and the medical charity sector. We are very grateful for these responses, as well as those we received from the following organisations:

- US Agency for Global Media
- Ranking Digital Rights
- Duke Reporters' Lab
- Internet Society India
- Pagella Politica/Facta.news
- Cognitive Security Collaborative Canada
- FairVote UK
- Tony Blair Institute
- Faktoje.al
- International Committee of the Red Cross
- Center for Countering Digital Hate
- Ofcom
- Institute for Strategic Dialogue
- Global Disinformation Index
- Media Policy Project, LSE
- Twitter
- Facebook
- BBC/Trusted News Initiative

- YouTube/Google
- Chequeado
- UK Government (Department for Digital, Culture, Media and Sport)
- MSI Reproductive Choices (formerly Marie Stopes International)
- Logically.ai
- Meedan

Glossary of terms

Claim clusters Clusters of claims that are related to each other, e.g. around a certain topic (such as Covid-19 vaccine side effects).

False narratives This phrase is used differently in different contexts, but here we are using it to refer to stories that connect and explain a set of events or experiences, which are formulated through news reports or online posts in multiple places and contain multiple false, misleading or only partially-correct claims and contribute to an inaccurate picture of a topic, event, institution or group of people. Here the emphasis is on what people end up believing as well as what is intended by e.g. activists, politicians or coordinated campaigns strategically disseminating information.

Harm The negative consequence(s) of a claim, narrative or information gap which affects individuals, groups, or institutions. In a public health context, this might be physical and immediate harm to individuals. In other contexts, this might mean losing money, reduced trust in institutions and decreased participation in democratic processes, or lack of compliance with public protection measures and advice put in place to protect society at large or specific communities.

Influence operations There are different interpretations of influence operations, but most encompass the following features: organised or coordinated efforts to manipulate or corrupt public debate or influence audiences for a strategic political or financial goal, often involving the perpetrator(s) concealing their identity via fake accounts or pages, and engaging in deceptive behavior.[10]

Information disorder and mis-/dis-/malinformation

- Misinformation is when false information is shared, but no harm is meant.
- Disinformation is when false information is knowingly shared to cause harm.
- Malinformation is when genuine information is shared to cause harm, often by moving information designed to stay private into the public sphere.[11]

Information incident A cluster or proliferation (sudden or slow-onset) of inaccurate or misleading claims and/or narratives related to and/or affecting perceptions of/behaviour towards a certain event/topic happening online or offline.

10 rand.org/topics/information-operations.html; carnegieendowment.org/2020/06/10/challenges-of-countering-influence-operations-pub-82031; about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf; rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c#page=17

11 rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c#page=17

References

Abdelaziz, Rowaida and Robins-Early, Nick, “How A Conspiracy Theory About The Notre Dame Cathedral Led To A Mosque Shooting”, HuffPost, 2019 huffingtonpost.co.uk/entry/bayonne-mosque-notre-dame-fire-conspiracy_n_5dc2fd22e4b0d8eb3c8e8a91

Cybersecurity & Infrastructure Security Agency, "CISA National Cyber Incident Scoring System", cisa.gov, as of March 2021

Donovan, Joan, “The Lifecycle of Media Manipulation”, The Verification Handbook 3, 2020 datajournalism.com/read/handbook/verification-3/investigating-disinformation-and-media-manipulation/the-lifecycle-of-media-manipulation

Digital, Culture, Media and Sport Committee, “Disinformation and ‘fake news’: Interim Report”, www.parliament.uk, 2018 publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/363.pdf#page=45

Funke, Daniel and Benkelman, Susan, “5 lessons from fact-checking the Notre Dame fire”, Poynter, 2019 poynter.org/fact-checking/2019/5-lessons-from-fact-checking-the-notre-dame-fire

HM Government, “Emergency Response and Recovery non statutory guidance accompanying the Civil Contingencies Act 2004”, gov.uk, 2013 assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/253488/Emergency_Response_and_Recovery_5th_edition_October_2013.pdf

Miller, Carl and Colliver, Chloe, “The 101 of Disinformation Detection”, The Institute for Strategic Dialogue, 2020 isdglobal.org/isd-publications/the-101-of-disinformation-detection

Nimmo, Ben, “The Breakout Scale: measuring the impact of influence operations”, Foreign Policy at Brookings, 2020 brookings.edu/wp-content/uploads/2020/09/Nimmo_influence_operations_PDF.pdf

Pamment, James, “The EU’s Role in Fighting Disinformation: Crafting A Disinformation Framework”, Carnegie Endowment for International Peace, 2020 carnegieendowment.org/2020/09/24/eu-s-role-in-fighting-disinformation-crafting-disinformation-framework-pub-82720

Rahman, Grace, “Here’s where those 5G and coronavirus conspiracy theories came from”, Full Fact, 2020 fullfact.org/online/5g-and-coronavirus-conspiracy-theories-came

Tran, Thi, Valecha, Rohit, Rad, Paul, Rao, Raghav, “Investigation of Misinformation Harms Related to Social Media During Humanitarian Crises”, University of Texas at San Antonio, 2020 researchgate.net/publication/339718919_An_Investigation_of_Misinformation_Harms_Related_to_Social_Media_During_Humanitarian_Crises

US Department of Defence, “Dictionary of Military and Associated Terms”, Joint Electronic Library, December 2020 jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf?ver=2018-09-28-100314-687

Wardle, Claire, “Fake news. It’s complicated”, First Draft, 2017 firstdraftnews.org/latest/fake-news-complicated

Wardle, Claire, and Derakhshan, Hossein, “Information Disorder: Toward an interdisciplinary framework for research and policy making”, Council of Europe, 2017 rm.coe.int/information-disorder-report-version-august-2018/16808c9c77

World Health Organisation, “Emergency Response Framework”, second edition, who.int, 2017 apps.who.int/iris/bitstream/handle/10665/258604/9789241512299-eng.pdf

YouGov, “YouGov / Today Programme Survey Results”, yougov.co.uk, 2016 d25d2506sfb94s.cloudfront.net/cumulus_uploads/document/x4iynd1mn7/TodayResults_160614_EUReferendum_W.pdf

Full Fact
2 Carlton Gardens
London
SW1Y 5AA

 fullfact.org/contact

 @FullFact

 fullfact.org

Published by Full Fact, October 2021.