MARCH 2021

A framework for information incidents

Consultation document

Closing Date for Responses: 14 May 2021

FULL FACT

About the framework for information incidents

Events such as elections, public health emergencies and natural disasters can affect the information environment in ways that make it harder to tackle misinformation effectively.

The responses from internet companies, governments, media, fact checkers, academics and civil society to Covid-19 misinformation in 2020 shows that those tackling misinformation can adapt and innovate fast. But these responses also highlighted the need for greater discussion of principles, proportionality, and the use of evidence in responding to other types of information incidents in future.

Full Fact is bringing together practitioners, experts and community groups from different sectors to develop a framework to identify how to respond to issues that occur during moments of crisis.

The aim of this consultation document is to test our thinking so far among a wide range of potential users. We extend our warmest thanks to those who contributed their time and gracious feedback throughout the first stage of this project, especially:

- 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

We refer to 'misinformation' throughout this document, but this framework is intended to also cover disinformation and malinformation as defined by Claire Wardle and Hossein Derakhshan:[1]

- Mis-information is when false information is shared, but no harm is meant.
- Dis-information is when false information is knowingly shared to cause harm.
- Mal-information is when genuine information is shared to cause harm (e.g. by moving information designed to stay private into the public sphere).

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, March 2021

Contents

Providing feedback

Summary of questions

Below is an overview of questions which are asked throughout this consultation document. Please answer as many questions as you want.

Page 9 covers a list of incidents which could cause significant changes to the information environment and create additional challenges for those tackling misinformation. The related questions are:

Q1. Is there anything missing, either as a category of information incident, or significant type of situation that might fall within a category?

Q2. Do you think that the distinction between categories of information incident is sufficiently clear? Yes/No. If No, please describe how this could be improved.

Page 12 sets out a system of five levels for determining an incident's severity. The related questions are:

Q3. Are these five levels of severity clear to understand and do the descriptors at each level of severity sufficiently cover the characteristics of information incidents? Yes / No. If No, please describe what needs to be clearer.

Q4. Which entities should be involved in assessing and declaring an incident's severity level in any country?

Q5. Can you describe (an) example(s) of what response(s) you/your organisation might introduce or look to see happening at different levels, for example at **Level 2**, **Level 3** and/or **Level 4**?

Page 20 sets out common challenges which arise across different types of incident. The related questions are:

Q6. Thinking about efforts to combat misinformation in exceptional circumstances, are there any important challenges missing or challenges that you would characterise differently?

Q7. Looking at the high level aims, are there any missing aims and/or significant responses which should be included here? (Please state the number of the challenge you are referring to.)

Page 26 pulls these components together. The related questions is:

Q8. Would this kind of tool be useful as a basis for discussions in your team and/or with other organisations?

Question 9 is for you to offer any other feedback:

Q9. Do you have any other comments on the framework (not covered by the previous questions)?

The last two questions are about you:

Q10. Please state the country you are based in and your name and email address.

Q11. If you work for an entity or organisation working on or affected by these issues, please state which one.

How to respond

There are two ways to respond to this consultation: either by email or via the Full Fact website. To respond via email, please send your answers to **policy@fullfact.org** (preferably in an attachment), indicating which question you are responding to.

Please feel free to answer as many or as few questions as you wish. To respond online, please fill in your answers at **fullfact.org/incidentframework**. Please note we will not be able to acknowledge every response.

You can respond to the ideas set out in this document by Friday 14 May 2021.

Following consultation, we will consider feedback on the framework's utility and main elements in order to refine the product content, before converting it into a useful tool and inviting organisations to test it in practice in the latter half of 2021.

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 creating 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.

In 2020 we saw 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. Full Fact brought together practitioners, experts and community groups from different sectors who are affected by and/or seek to affect the information environment to develop this framework. We sought to identify:

- A methodology for assessing the severity of an information incident
- The most common issues that occur during information incidents
- Joint aims for how organisations should try to respond to these issues

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 are producing 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 response guidance from public health bodies and governments.[1]

In developing this work we have drawn on existing research and analysis from First Draft, the Carnegie Endowment for International Peace, Ben Nimmo, Joan Donovan, the University of Texas researchers and others to develop a methodology for determining and describing the severity of different information incidents. We are grateful to the authors of these reports for laying the groundwork to understand these complex issues.

We would now like to test this thinking among a wider range of potential users, in order to develop a simple and useful tool to help specialists in this area coordinate work, and to help other stakeholders better understand and engage with that work. In this consultation, we are particularly keen to hear from organisations, groups and individuals that identify, analyse and respond to misinformation, and those whose work, audiences and/or service users are affected by misinformation.

You can respond to the ideas set out in this document by Friday 14 May 2021.

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

What is an information incident?

Our aim is to identify the types of incidents that are likely to have a substantial and material impact on the people, organisations and systems that consume, process, share or act on information – referred to in this document as the "information environment".[2] Importantly, the information environment informs decisions – whether by citizens, other actors or policy-makers – 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 or referendum (particularly where referendums are not normal occurrences) 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, we propose nine categories of information incidents that might require responses above and beyond 'business as normal'. These categories are non-exhaustive and are broad by design. They are intended to give an understanding of the types of situations where one could reasonably expect the information environment to be affected. Not all incidents will require the same level of response: later we present a methodology for determining the severity of an incident, and ways to translate this into proportionate responses.

Categories of incidents:

- **Human rights or freedom of expression abuse.** Civil action such as disrupted peaceful protests; violent public confrontations; long term escalating tension e.g. between regions; mass detainment and/or killings; citizenship or demographic changes.

[2] This description of "the information environment" is adapted for the purpose of discussing largely domestic misinformation from a definition used by the military in relation to the hybrid warfare context.

- **Unexpected disasters with high, wide reaching impact.** Deliberate attack with wide reaching, long-term impact; innovative or novel deliberate attack; incidents without warning that create deaths or displacement of people; national/regional health incidents; natural disasters affecting people.
- **Unexpected events where response plans are likely to exist, with low and/or short term impact.** High impact weather; incidents with cultural/religious significance; localised deliberate attacks with short term impact; accidents and transport disasters; some hacks, leaks or data dumps. Authorities would be expected to have a plan for response or containment for these types of incidents.
- **Long-horizon and long-tail incidents.** Economic changes; shortages e.g. food or fuel; exceptional electoral events; displacement of people; climate change.
- **Planned political events in democratic states.** National or regional votes.
- **Nationally significant events where there could be opportunities to exploit polarisation.** War memorials; societal or political commemorations; controversial political incidents.
- **High impact incidents that occur or where the impact is felt across multiple markets.** Pandemic or war.
- **Controversial and/or shocking news stories.** Events that generate news headlines, but do not become major incidents.
- **Spreading of false/misleading information by authoritative actors.** Some party political messages; high profile endorsements of conspiracies; denying controversial incidents.

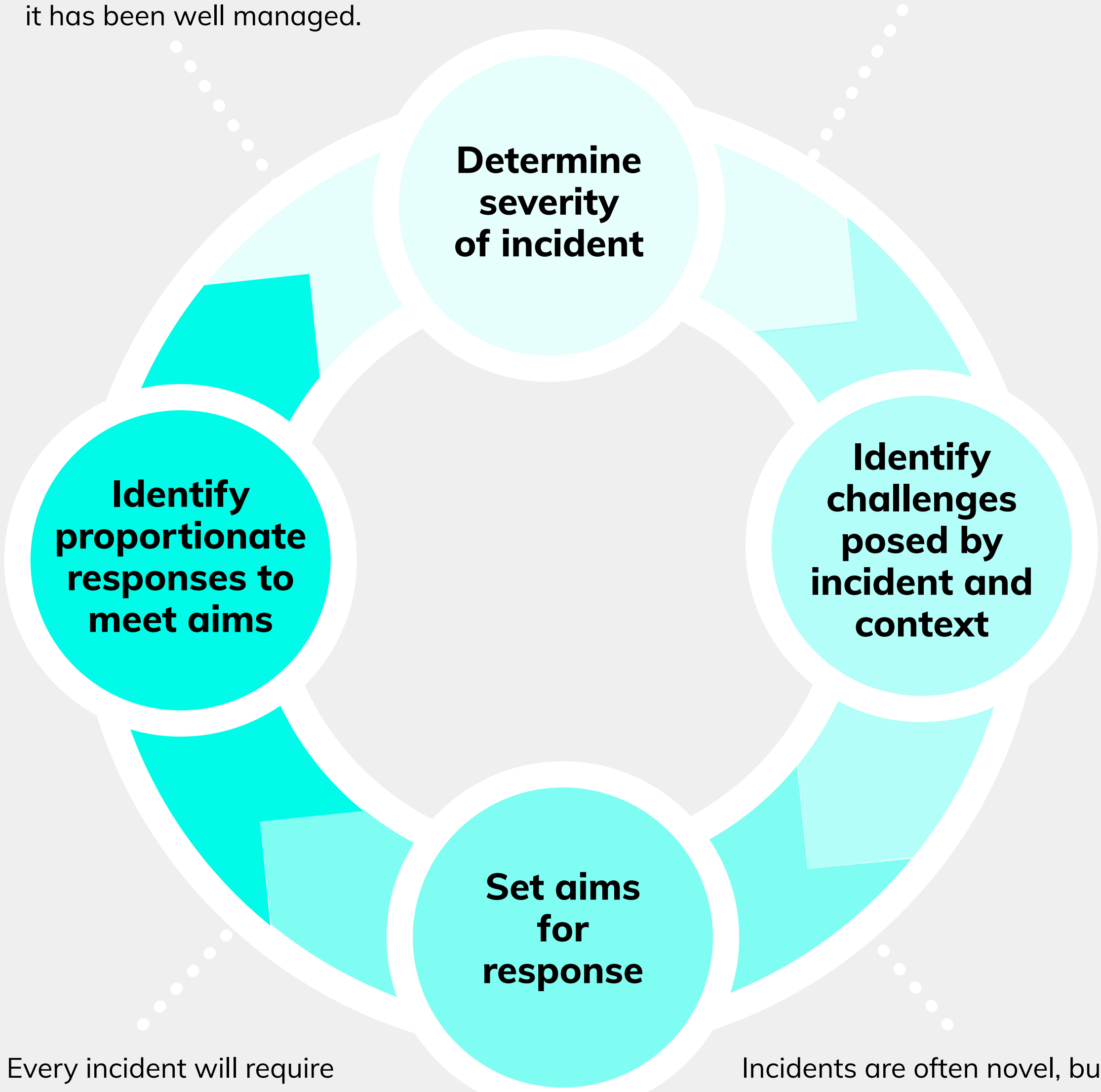
Q1 *Is there anything missing, either as a category of information incident, or significant type of situation that might fall within a category?*

Q2 *Do you think that the distinction between categories of information incident is sufficiently clear? Yes/No. If No, please describe how this could be improved.*

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.



Determine severity of incident

Identify challenges posed by incident and context

Set aims for response

Identify proportionate responses to meet aims

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.

Five levels of severity

Beyond these broad categories of incidents in scope we wanted to understand which incidents were more severe than others, and whether there were common attributes that could distinguish between them.

The below methodology attempts to do this, with the goal being that different levels will require different actions to keep the reponses proportionate, and which can be ramped up or lowered as needed. We propose five levels to map this 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, but where organisations have space to focus resources on long-term goals and priorities.

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” that indicate when an incident can be moved down in severity will be a key priority for the next phase of developing this framework.

Levels of incidents

Level 1 ● ○ ○ ○ ○

Business as normal, no additional response needed

Reach: Misinformation is circulating at levels considered to be normal, with no incident apparent on the horizon.

Subject matter: There may be some spikes of claims around certain events, topics or locations but they die down quickly.

Urgency of response: There is space to work on long-term goals such as audience resilience and media literacy.

Collaboration: Key organisations are collaborating to identify emerging threats.

Level 2 ● ● ○ ○ ○

Monitor and prepare external facing responses

Reach: False claims or narratives are breaking out related to a certain topic or event, on one or more smaller platforms.

Urgency of response: Subject matter may be controversial or polarising, or has potential to cause harm if the volume of misinformation or engagement grows, e.g. may threaten vulnerable groups.

Collaboration: There is time to put plans in place to mitigate the growth and effects of misinformation.

Level 3 ● ● ● ○ ○

An incident is occurring, responses ramp up

Reach: Misinformation is moving into the mainstream and there may be amplification (whether organic or coordinated).

Continues...

Urgency of response: The incident may be relatively short term, so acting immediately is important.

Collaboration: Individual organisations are implementing internal incident response plans; collaboration is enhanced to identify cross-media and cross-platform risks.

Level 4 ●●●●○

An incident is occurring, coordinated responses start

Reach: Misinformation is moving into the mainstream / being amplified by large accounts; there may be evidence of coordination or malicious intent or unprecedented use of online campaigning techniques/technology.

Subject matter: The event or topics may be completely new, or emotive, to the audience, increasing the likelihood of confusion; the topic(s) may be polarising.

Urgency of response: The misinformation might spark violence or physical danger; the incident may compound existing conspiracy theories / false narratives.

Collaboration: Core and expanded (e.g. public health, agriculture) organisations are collaborating.

Level 5 ●●●●●

Maximum response levels and co-operation required

Reach: The incident is unusual in terms of fast spread and high volumes of misinformation, and is unlikely to be resolved in the short term; the incident is global or affecting multiple regions, with the same misinformation often appearing in different languages.

Urgency of response: The misinformation is causing, or is likely to cause, significant human harm; lasting effects on public misperceptions may occur.

Collaboration: A maximum level of cooperation from a range of different organisations is required.

Deciding on a level

One of the reasons to have a shared framework is to improve coordination and cohesion and establish shared language. From that perspective, it would be beneficial for different organisations to agree the level of an incident together. 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).

An approach where each organisation reaches their own determination allows for flexibility, but may become inoperable. We recognise the difficulty for large multinational companies and governments to commit to something too rigid, but equally recognise the tension in creating something that cannot achieve the aims set out above, including enabling quick responses. We welcome views on which approach would function best.

Impact of levels

The purpose of having different levels is to enable greater understanding and to identify what a proportionate response would be. 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. Incidents may not exhibit every single feature described at each level above.

We seek to define this further below but welcome views on what types of measures you would expect to occur at each level.

Q3 *Are these five levels of severity clear to understand and do the descriptors at each level of severity sufficiently cover the characteristics of information incidents? Yes / No. If No, please describe what needs to be clearer.*

Continues...

Q5 *Which entities should be involved in assessing and declaring an incident's severity level in any country?*

Q6 *Can you describe (an) example(s) of what response(s) you/your organisation might introduce, or would want to see happening at different levels, for example at **Level 2**, **Level 3** and/or **Level 4**?*

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.[3] Level 2 characteristics included:

- False claims and narratives breaking out about safety of 5G on one or more platforms
- Potential for real world harm in the future, but no immediate threat
- Time to put plans in place to mitigate growth of misinformation

Public health information about the safety of 5G rollout was not improved at the time when Full Fact identified this emerging incident. 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.

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

[3] Rahman, Grace, “Here’s where those 5G and coronavirus conspiracy theories came from”, Full Fact, April 2020

Vest protester was seen in a tower.[4] 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:

- Unexpected event, where acting immediately to respond to misinformation is important.
- Marginalised groups more likely to be targeted with misinformation.
- Misinformation breaking out onto mainstream platforms and the incident attracting news coverage.

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[5].

Level 4: Brexit Referendum, UK, June 2016

Fresh from a recent election victory in 2015, the UK government announced a referendum on the UK's membership of the EU. One campaign ran a billion targeted Facebook ads (a lot, at the time), while the other campaign was criticised for using government funds to distribute partisan leaflets. Evidence of malicious coordination was proved retrospectively[6], but other **Level 4** indicators present at the time included:

- Unfamiliar event (a UK referendum) with complex and hard-to-communicate topics (e.g. EU law and trade).
- Polarising campaign with misleading claims amplified by high profile/large accounts and compounding existing false narratives (e.g. 10%/70% of UK law comes from the EU).

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

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

[6] Digital, Culture, Media and Sport Committee, "Disinformation and 'fake news': Interim Report", parliament.uk, 2018

- Unprecedented use of technology to distribute unscrutinised claims to certain groups.

At the time the response to these high levels of misinformation was limited to fact checking, which – despite best efforts – struggled to reach a large proportion of the UK population who did not trust the experts attempting to inform them.[7]

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 volume of misinformation. The characteristics that put this incident at the most severe level include:

- Unusually high volumes and fast spread of life threatening misinformation on multiple platforms.
- Extended time period, with lasting effects on public misperceptions and public health.
- Unprecedented levels of cooperation among different organisations to effectively combat the incident.

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.

[7] For example, a YouGov poll carried out in June 2016 found that trust in economists stood at 38% and trust in academics stood at 42% ("YouGov / Today Programme Survey Results", yougov.co.uk, 2016)

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. We have seen many incidents that threaten freedom of expression, create an unclear or quickly changing situation, or create longer lasting problems that outlast the initial incident, such as conspiracy theories or 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. In this section, we present a set of aims for responding to the most common challenges we have identified. This is likely to be non-comprehensive, and we would welcome views on whether there is anything significant missing.

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 e.g. when there is:

- 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
- Suspected or known foreign interference

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
- Enable independent experts to scrutinise AI recommendations

2. Unclear or quickly changing situation

Challenges e.g. when there is:

- Lack of insight into type and scope of misinformation and/or movement of content between platforms
- Unhelpful duplication of efforts among organisations
- Contradictory interpretations of a situation

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

- Share monitoring and verification information between trusted experts
- 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

Continues...

3. Difficulty disseminating or communicating information

Challenges e.g. when there is:

- Low baseline knowledge of key issues among public, politicians and media
- Low statistical literacy among public and media
- Accurate information is not contextualised or adapted for certain groups
- Topics are complicated or highly technical
- Information overload and audiences 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

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

- Promote relevant impartial or official sources of information
- Identify and engage with appropriate trusted voices to disseminate information
- Disseminate information to pre-empt belief in emerging conspiracy theories

4. Information vacuums and uncertainty

Challenges e.g. 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 reliable information from authoritative sources is available and that any limitations are communicated

- Funding and resources for statistical offices and impartial information providers

Continues...

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

5. Unhelpful behaviour by influential public figures

Challenges e.g. when high profile figures:

- Repeat false claims or make conflicting statements
- Cast doubt on accurate information
- 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. Pressure to work at speed and scale to halt spread of false beliefs

Challenges e.g. 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

Continues...

Aims: Limit bad information and ensure corrective information appears when and where people need it, and have 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

7. Immediate threats to public order and safety

Challenges e.g. when:

- □ Public order and safety is dependent on the public understanding information accurately
- □ Communication from affected communities and first responders is compromised or ignored
- □ False information creates potential for physical harm through violence or hazard

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

- □ 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 an incident or incidents

Challenges e.g. when:

- □ The incident spawns or entrenches conspiracy theories or myths which outlast the incident

Continues...

- False narratives are repeated over years and create hard-to-shift public misperceptions

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

Q7 *Thinking about efforts to combat misinformation in situations above* ***Level 1****, are there any important challenges missing or challenges that you would characterise differently?*

Q8 *Looking at the high level aims, are there any missing aims and/or significant responses which should be included here? (Please state the number of the challenge you are referring to.)*

How do severity levels map across to challenges and responses?

After this consultation, we aim to bring these elements together to create a simple, coherent product that can be consulted when incidents occur and be a common reference tool for effective action.

We want to match up the elements outlined above: understanding the severity of an incident and mapping out common 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.

Below we outline 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.

Worked example: information vacuums and uncertainty

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

Responses that could meet this aim **in a Level 2 scenario**, when false claims are breaking out on smaller platforms, which may be polarising or threaten vulnerable groups:

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

Responses that could meet this aim **in a Level 3 scenario**, 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

Continues...

Responses that could meet this this aim **in a Level 4 scenario**, when misinformation is being amplified, there may be evidence of coordination, and there is potential for violence 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

Responses that could meet this aim **in a Level 5 scenario**, 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 straight away to improve responses
- Inter- and cross-industry collaboration to create new policies and products to respond to the situation

Following feedback from consultation, we intend to create a tool to articulate the different response levels for the other challenges.

Q9 *Would this kind of tool be useful as a basis for discussions in your team and/or with other organisations?*

Conclusion

Finally, thank you for your interest in this work. Please do submit any views on the questions outlined in this consultation or any further views. The consultation will be open until Friday 14 May 2021 to allow time for your consideration and response.

There are two ways to respond to this consultation: either by email or online via the Full Fact website. If you are providing your input by email please simply send your answers to **policy@fullfact.org** (preferably in an attachment) and indicate which questions you are responding to. Please feel free to answer as many or as few questions as you wish. If you would like to respond online, go to the website at **fullfact.org/incidentframework** and fill in your answers there.

We look forward to hearing from you.

Annex: Further 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.

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, March 2021.