



Managing Misinformation in the AI Age

hello@helpfuldigital.com

Who we are



Tim Lloyd
Managing Director

- ▶ Local and trade journalist then a Head of Digital Communications for the UK government.
- ▶ Led social media strategy through issues, including elections, industrial unrest and health epidemics.



Claire Turner
Associate Director

- ▶ Working at the forefront of social media and crisis comms for over a decade.
- ▶ Helps some of the world's largest brands to develop their crisis communication skills.

Who we are



Sophie Pattison
Digital Account Executive



Claire Turner
Associate Director

Today's agenda

Best practices: Understanding modern misinformation (40mins)

Quick break (10 mins)

Introduction to the platform, Simulation Phase 1: Strategy Development, and Phase 1 Debrief (65 mins)

Quick break (10 mins)

Best practices: Messaging to Maintain Trust (10 mins)

Simulation Phase 2: Messaging, channels and resources and Phase 2 Debrief (40 mins)

Final thoughts (5 mins)

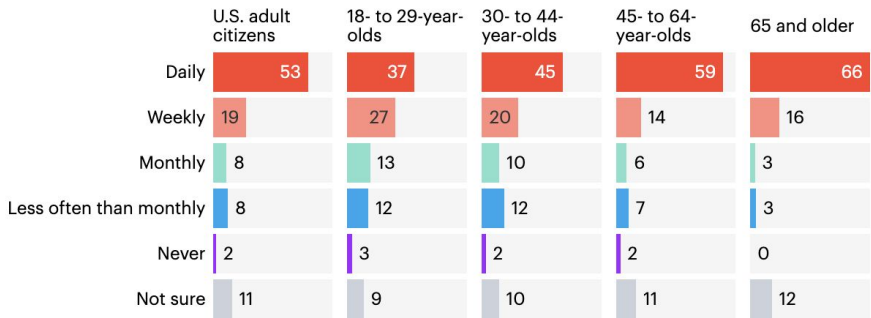


Session objectives

- ▶ Understand the full spectrum of misinformation and the motivations behind it
- ▶ Recognize how AI is reshaping misinformation
- ▶ Practical verification techniques
- ▶ Strengthen ability to communicate clearly and credibly under pressure
- ▶ Practise coordination and decision-making via an immersive, scenario-based tabletop exercise.

Half of Americans say they see what they think is false or misleading information online **every day**

How often do you see what you think is false or misleading information online? (%)



YouGov

April 3 - 9, 2023

How has misinformation impacted your work?

Understanding Misinformation

Types of misinformation

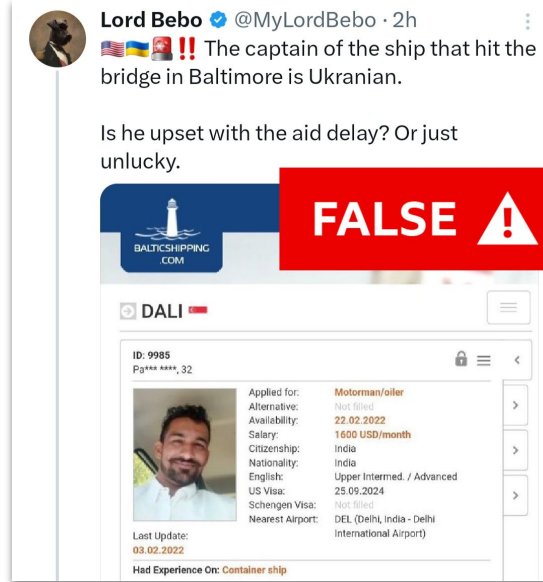
Misinformation

Not created or shared with the intent to cause harm



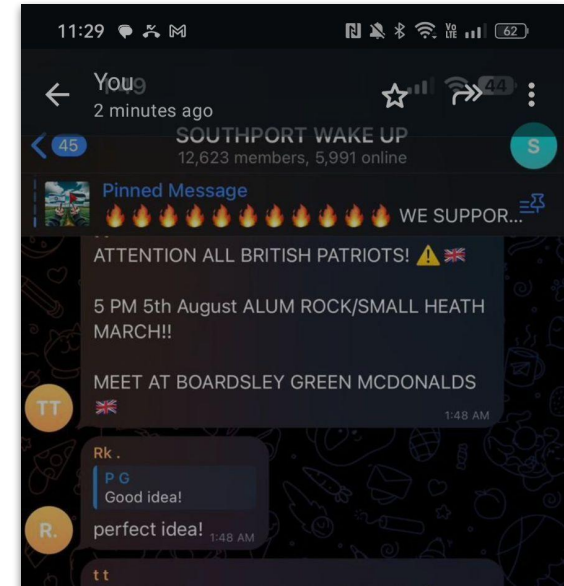
Malinformation

Based on fact, but used out of context to mislead, harm, or manipulate

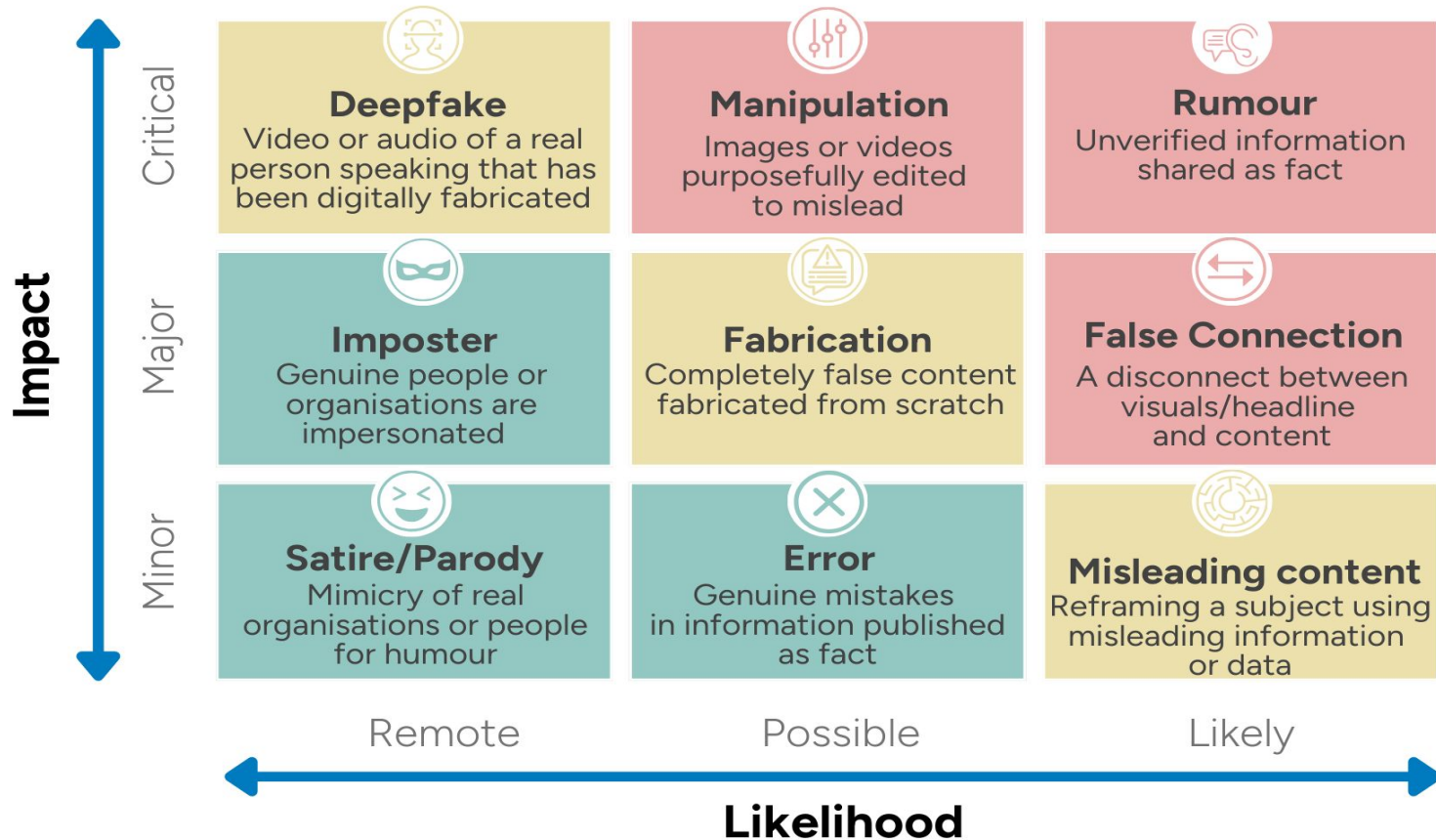


Disinformation

Intended to mislead, harm, or manipulate



Types of **mis**- and **dis**information



Why do we believe it?



Confirmation bias



Social media design



Attention and
impatience



Illusory truth effect



Emotional barriers



Inevitable data
vacuum

The role of the algorithm

- ▶ Content suggestions
- ▶ Illusory truth effect
- ▶ Faulty search suggestions



How does it spread?



Humans

Mostly innocent

- ▶ Sharing becomes a habit
- ▶ Misinformation superspreaders
- ▶ Feeds the algorithm

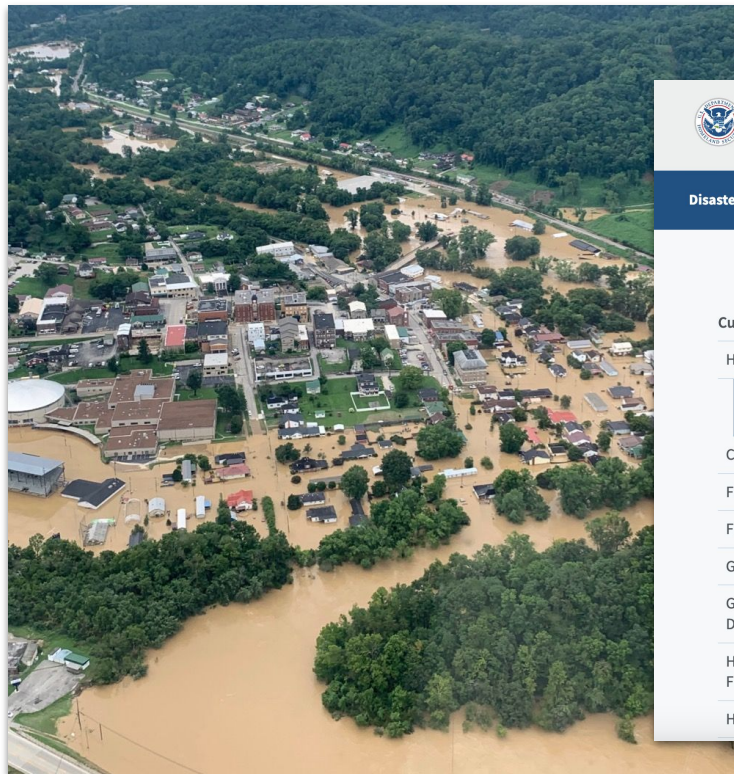



Bots


Usually bad actors

- ▶ Utilize various technology, including AI
- ▶ Artificial amplification
- ▶ Plays on weaknesses

Growing risks



 **FEMA**

Search FEMA.gov 

Disasters & Assistance ▾ Grants ▾ Floods & Maps ▾ Emergency Management ▾ About ▾ Work With Us ▾ [Apply for Assistance](#)

Current Disasters

Hurricane Helene

Hurricane Rumor Response

Connecticut Severe Storm

Florida Hurricane Debby

Florida Hurricane Helene

Georgia Hurricane Helene

Georgia Tropical Storm Debby

Hermit's Peak/Calf Canyon Fire

Hurricane Milton

Hurricane Rumor Response

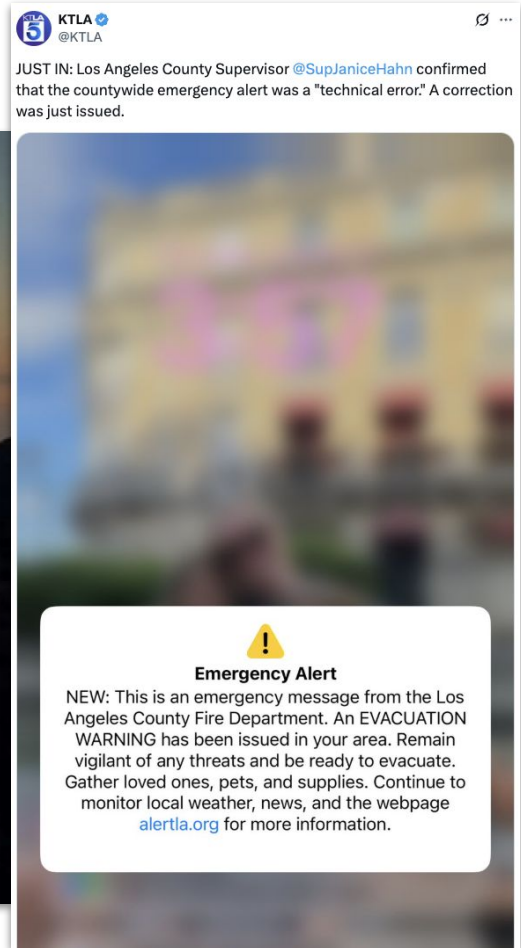
العربية Deutsch English Español Français Kreyòl 한국어 Burmese
Português, Brasil Русский Tagalog 简体中文

Help keep yourself, your family and your community safe by being aware of rumors and scams and sharing [official information from trusted sources](#).

Do your part to stop the spread of rumors by doing three easy things:

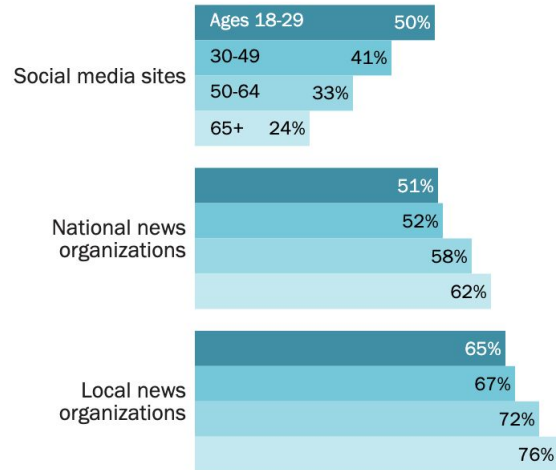
1. Find trusted sources of information.
2. Share information from trusted sources.
3. Discourage others from sharing information from unverified sources.

Growing risks



Younger adults trust info from social media sites about as much as from national news organizations

% of U.S. adults who say they have **a lot of** or **some** trust in the information they get from ...

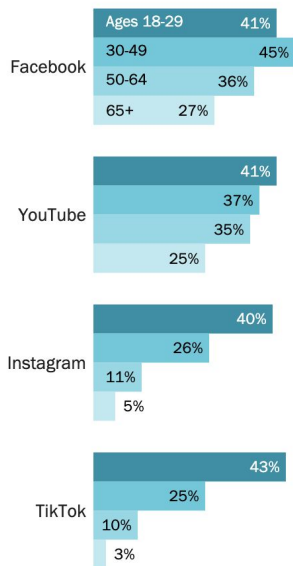


How do these dynamics show up in government communications?

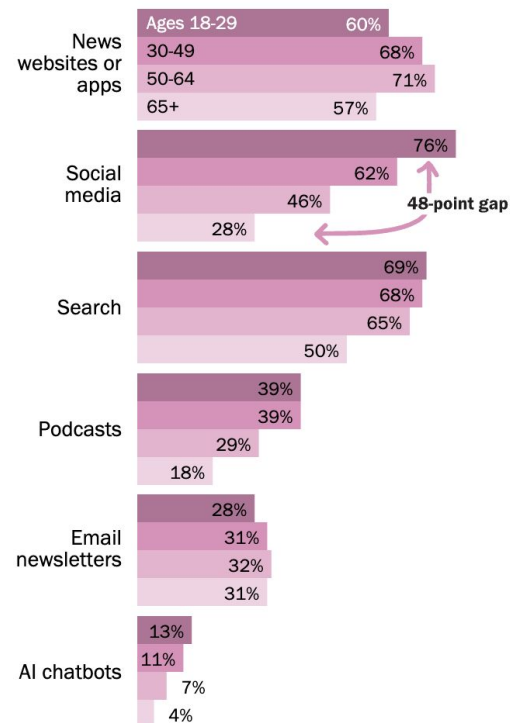
Latest news consumption trends

News consumption on social media sites, by age

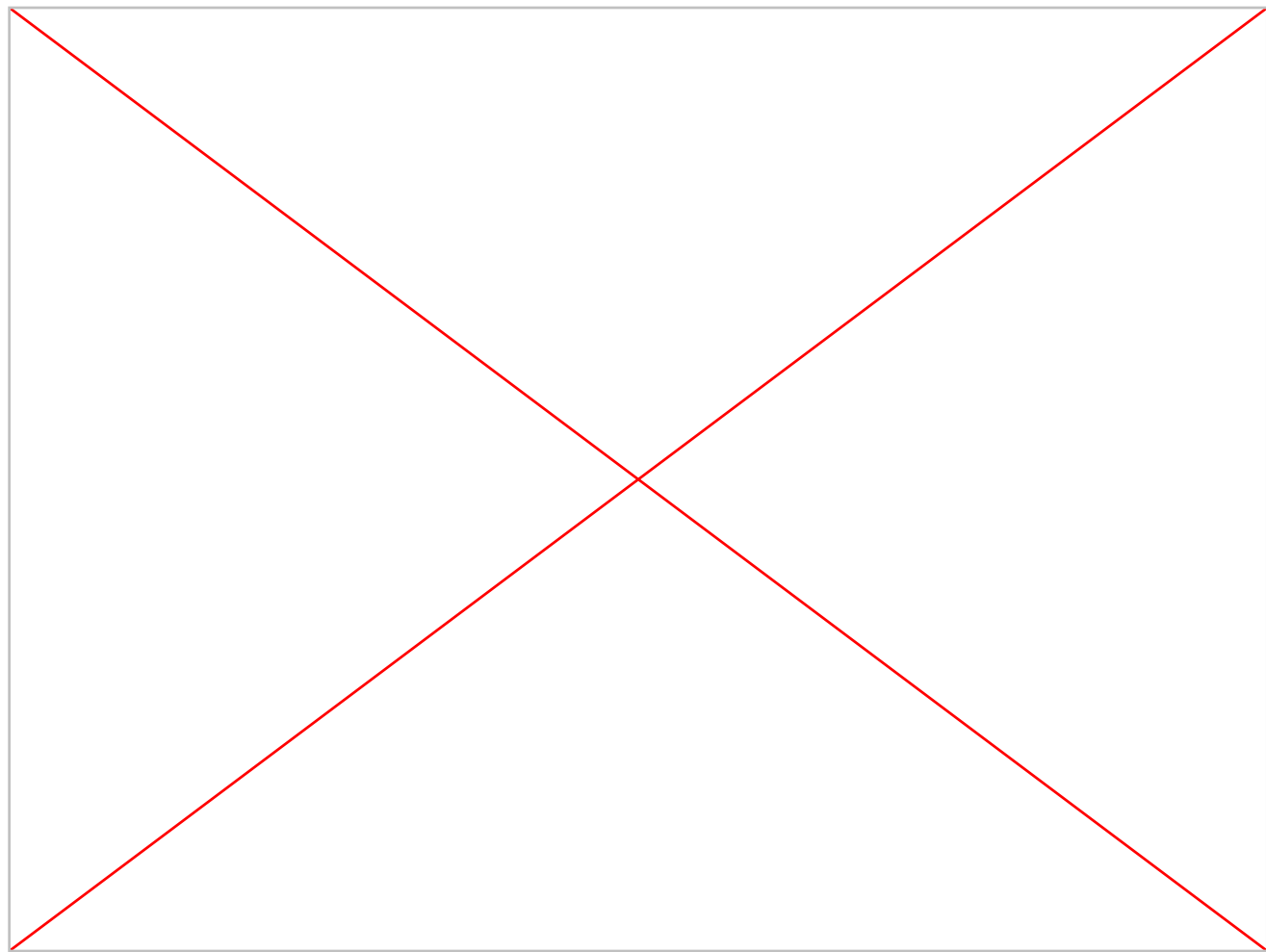
% of U.S. adults who say they **regularly** get news on each social media site



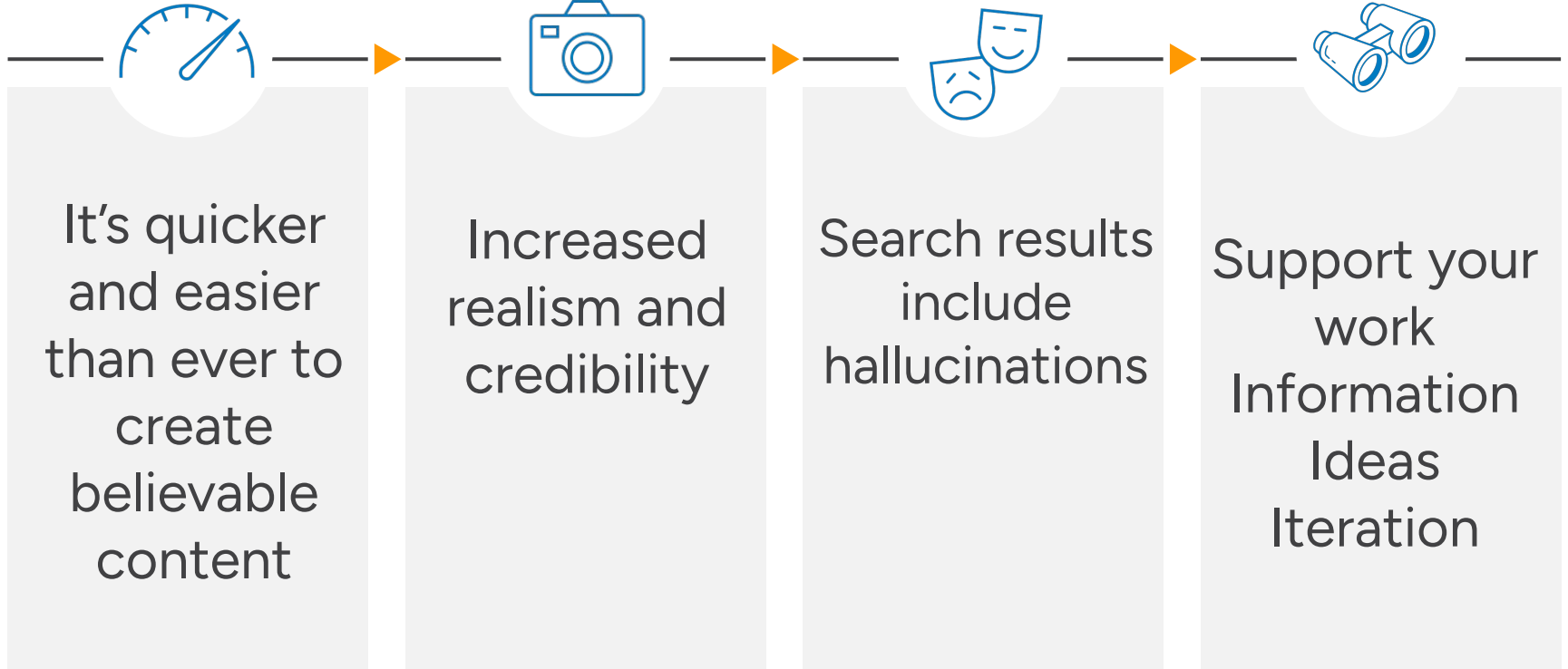
% of U.S. adults who say they **often** or **sometimes** get news from ...



The AI Effect: New Tools; New Threats



AI has changed the landscape




Hallucinations

[All](#) [News](#) [Images](#) [Videos](#) [Short videos](#) [Forums](#) [Shopping](#) [More](#) [Tools](#)

AI Overview



The most recent fatal Airbus crash occurred on **Thursday, June 12, 2025**, in India, involving an Air India flight. The plane crashed shortly after takeoff in Ahmedabad, killing a significant number of passengers and crew members. The crash involved an Airbus A330-243, bound for London, with 242 people on board. [🔗](#)



Key Details:


- **Date:** Thursday, June 12, 2025
- **Location:** Ahmedabad, India
- **Aircraft:** Airbus A330-243
- **Airline:** Air India
- **Fatalities:** Significant loss of life
- **Passengers and Crew:** 242 on board [🔗](#)


AI responses may include mistakes. [Learn more](#)



India Plane Crash Live Updates: Air India Flight With 242 ...


15 minutes ago — An Air India flight carrying 242 passengers and crew members crashed...


 The New York Times [⋮](#)



'Many killed' as Air India plane with 242 people on board ...

45 minutes ago — Follow our live coverage here. NEW DELHI — An Air India plane bound...

 The Straits Times [⋮](#)



Identify and Interrogate

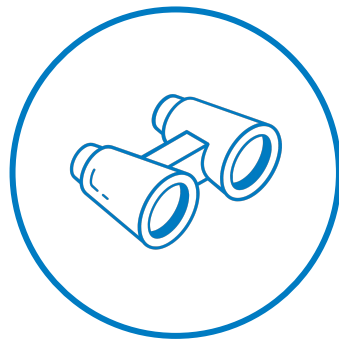
SIFT to identify misinformation



Stop



Investigate



Find better
coverage



Trace

How to identify and verify



TinEye



Google Lens

Google Fact Check Tools

AFP Fact Check

Strategies for countering misinformation

The importance of monitoring

×

Advanced search

Search

Words

All of these words

wildfire

Example: what's happening · contains both "what's" and "happening"

This exact phrase

Example: happy hour · contains the exact phrase "happy hour"

Any of these words

New Jersey NJ

Example: cats dogs · contains either "cats" or "dogs" (or both)

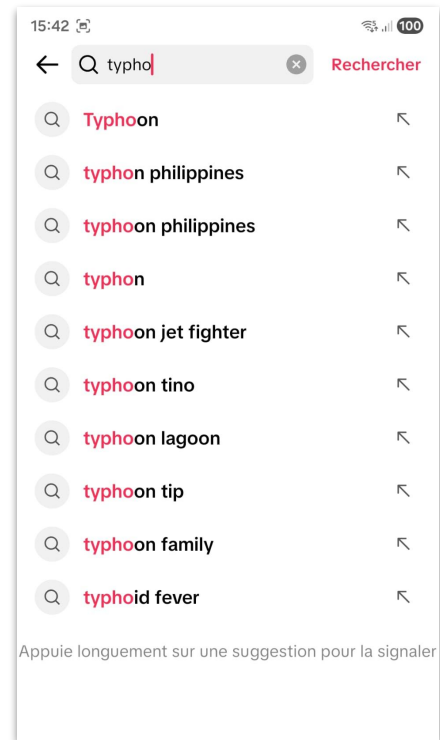
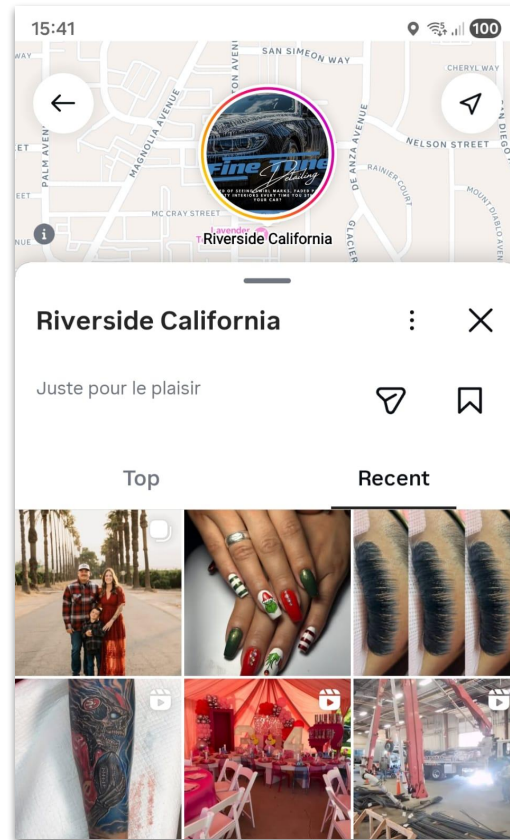
None of these words

Example: cats dogs · does not contain "cats" and does not contain "dogs"

These hashtags

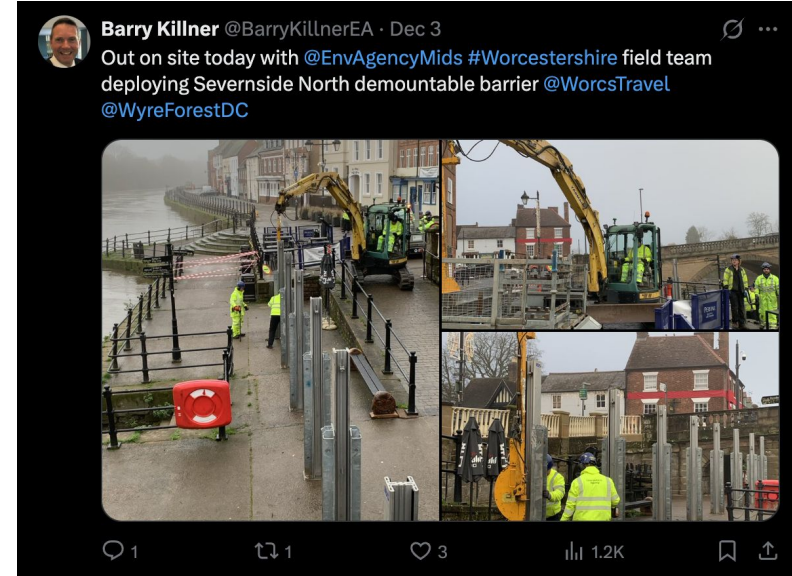
Example: #ThrowbackThursday · contains the hashtag #ThrowbackThursday

Language



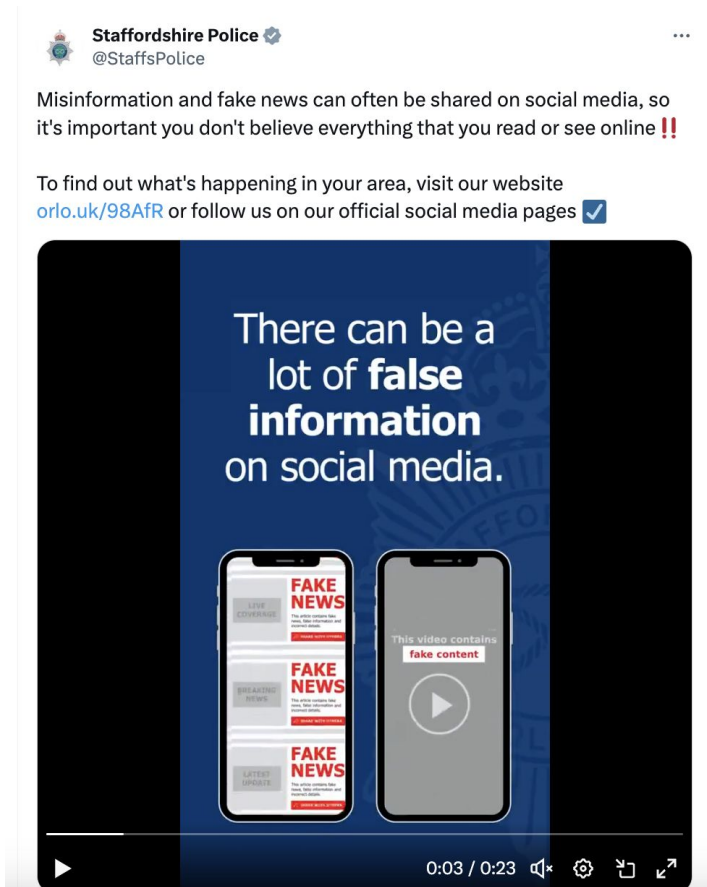
Getting ahead

- ▶ Stay authentic and up to date
- ▶ Utilize trusted voices
- ▶ Knowledge is power











Prime your audience

- ▶ How to spot misinformation
- ▶ Promoting alert channels
- ▶ What happens in an emergency



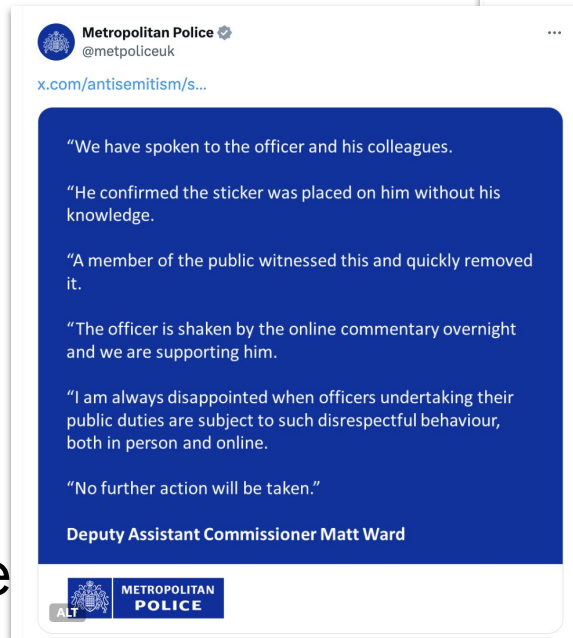
Misinformation considerations

	Does this information affect our ability to deliver services?
	Does this information impact the people who depend on our services?
	Does this information pose a significant risk to the public?
	Is it clear that it's a parody, joke, or meme?

	Is the person spreading misinformation where anyone will see it?
	Is the subject easy to understand?
	Is the misinformer promoting fake experts?
	Should the misinformer know better?

First steps for handling

1. Pause. Think. Consider.
2. Verify the information
3. Escalate as needed
4. Respond
 - a. Individual vs aggregate response
 - b. Be strategic



Today in London: @metpoliceuk officer with "🇮🇱 Boycott Israeli apartheid" sticker.

Britain 🇬🇧 has fallen.

@AC_Black_ @DrEliDavid



Readers added context to this image on another post

The sticker was placed on this police officer's uniform without his knowledge.

A member of the public witnessed this, and quickly removed it.

x.com/metpoliceuk/st...

Maintain credibility and confidence

Explain falsehoods or inaccuracies

Reply to superspreaders

Link to evidence or other sources

Avoid debates

Be clear and succinct

Encourage sharing

How to respond to video misinformation

- ▶ Identify the audience affected
- ▶ Choose a spokesperson
- ▶ Respond in-platform
- ▶ Respond on alternative platforms



Break

Crisis90 Phase 1

Setting up for success

- Agreeing roles and responsibilities – who is doing what?
- Agreeing on priorities and working your process!
- Establishing the right rhythm of meetings (Meet > **Break** > Meet)
- Identifying information gaps, logging latest information, and sharing information amongst teams/agencies

Agree roles and responsibilities

Social media responder

- Draft updates that cover key actions and information
- Adapting messaging to appropriate platforms
- Triage/responding to comments and questions

Social media monitor

- Monitor traditional and social media using keywords and search term
- Consistently compile and share out monitoring reports

Field PIO

- Identify staging areas
- Communicate with on-site incident command and the broader JIC for messaging

Remember to stay strategic

Using a **'people first'** strategy to serve as your guiding star for what you will **do** and what you will **say**:



Residents



Responders

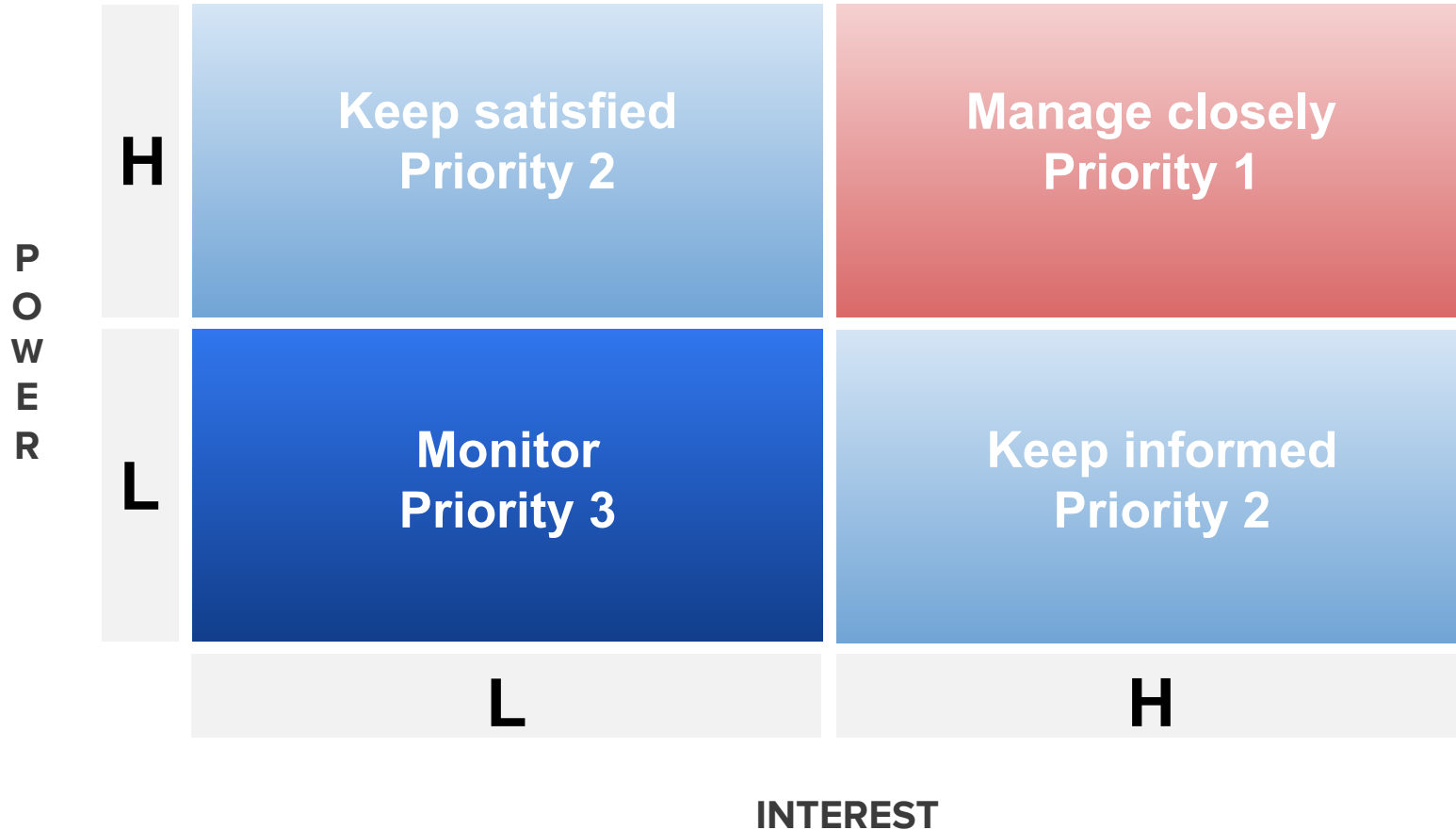


Staff



Public

Prioritization: stakeholder mapping



Where to reach your audience

Who?



Politicians



Journalists



Staff



Public

What?

Key info

Action plan

Where?

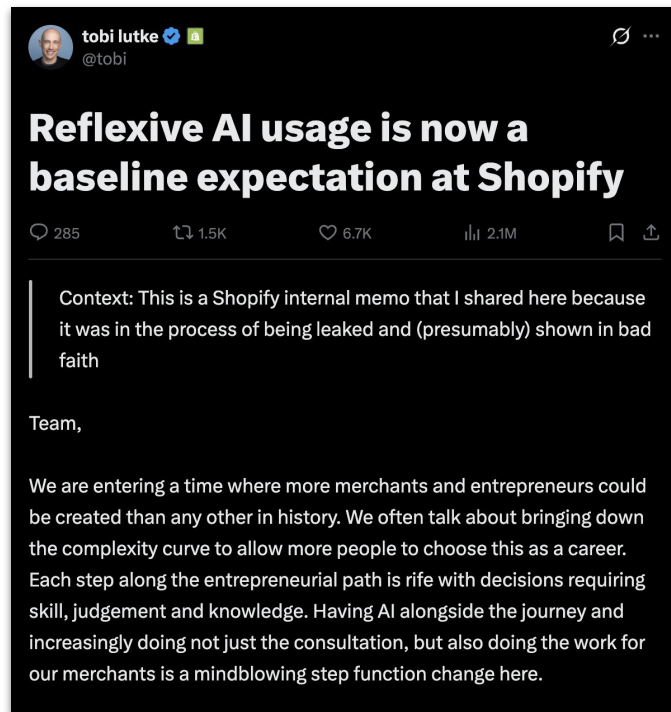


Internal communications

Don't forget internal comms networks too.

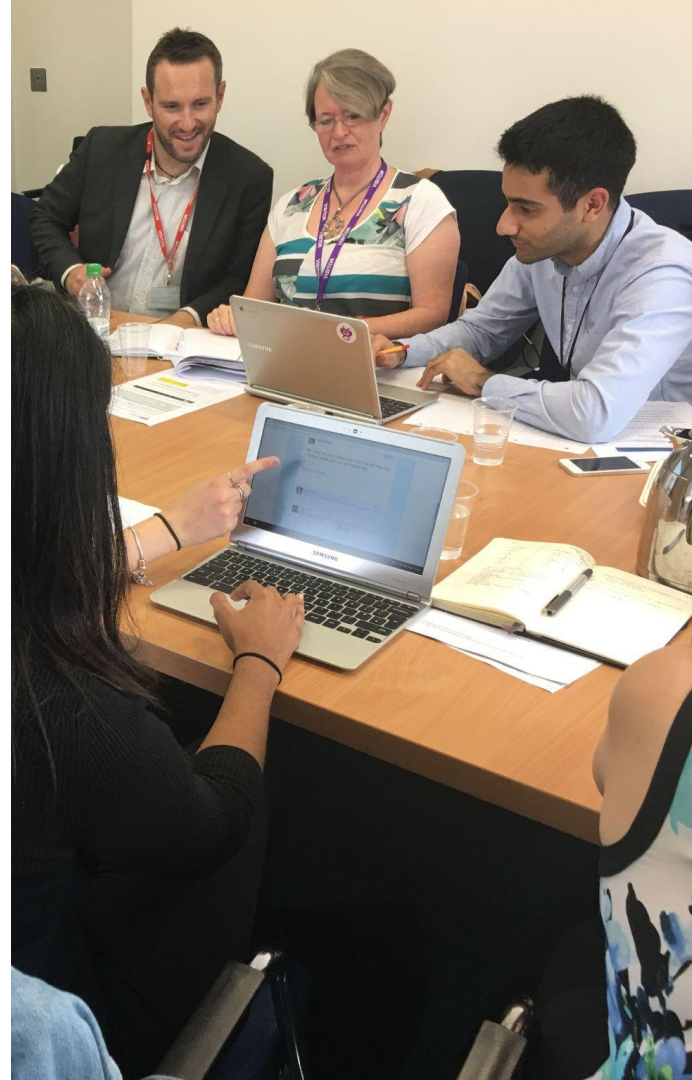
Avoid people going rogue with rumors or speculation.

Remember: Any information posted internally will likely be seen outside the organization.



The training

- 2 phases of an escalating crisis
- You are members of the government communications team for Riverside City Government.
- Working in teams to complete tasks within tight deadlines
- Appoint a scribe and a spokesperson for your team



Guidelines

- Don't challenge the scenario...
- You are the team responsible for leading the response
- Teamwork makes the dream work!
- Draw on your expertise and experience to complete the tasks

Scenario

- A Category 4 hurricane has recently impacted multiple towns across the region of Riverside, causing widespread flooding, property damage and disruption to essential services. In the hours following the storm, misinformation is spreading rapidly across social media and other channels, complicating your team's ability to communicate effectively with the public.
- It is your team's responsibility to start to counteract the misinformation circulating online to ensure that accurate information is provided to members of the public and local stakeholders.

Accessing the platform

smsworkshop25.crisis90.com

Username: blue
Password: sky265dec

Username: green
Password: grass398dec

Username: yellow
Password: dress763dec

Username: red
Password: rose852dec

Username: orange
Password: ball418dec

Username: purple
Password: paint863dec

Accessing the platform

smsworkshop25.crisis90.com

Username: brown
Password: cocoa24

Username: pink
Password: princess24

Username: black
Password: bear24

Username: grey
Password: goose24

Phase 1: Debrief

Strategy

People first! Note key impacts to public and inform stakeholders of these

Stakeholders

1-5: Local politicians, residents, external public, FEMA, journalists

Managing Misinformation

Social media responder; media monitor; field PIO

Priority actions

Assemble team and assign roles, establish situational awareness, identify immediate harms, agree key messaging points, locate and distribute resources

Break



Click to add title

Messaging and resources to maintain trust

Warn and Inform

Warn

Warn stakeholders of the present dangers, like wildfires, floods, impending storms, carbon monoxide poisoning, etc.

Inform

Inform them of what key actions they should be taking, like evacuation, preparing for a storm, closing windows, etc. Don't forget to include what channels they should be monitoring for more information.

A Call to Action is key

**NWS GSP** 
@NWSGSP

Follow

URGENT: FLASH FLOOD EMERGENCY FOR THE LAKE LURE DAM! DAM FAILURE IMMINENT! RESIDENTS BELOW THE DAM NEED TO EVACUATE TO HIGHER GROUND IMMEDIATELY! [#ncwx](#)

**NWS GSP**  @NWSGSP · 47m
Flash Flood Emergency continues for Rutherford County, NC until 2:00 AM EDT

Flash Flood Emergency

Valid Until
2:00 AM EDT Saturday
September 28, 2024

Safety Information

 Move immediately to higher ground!

 Avoid walking or driving through flood waters!

Potential Exposure

Population: 3,052
Schools: 0
Hospitals: 0


ALT

 This is a life threatening situation. Seek higher ground now!



This is a matter of life or death

Tone and language

- Keep the tone confident, serious and reassuring.
- Avoid technical language and acronyms
- Use inclusive, simple language
- Break up large chunks of text into shorter sentences and paragraphs
- Too much detail can actually hinder rather than help

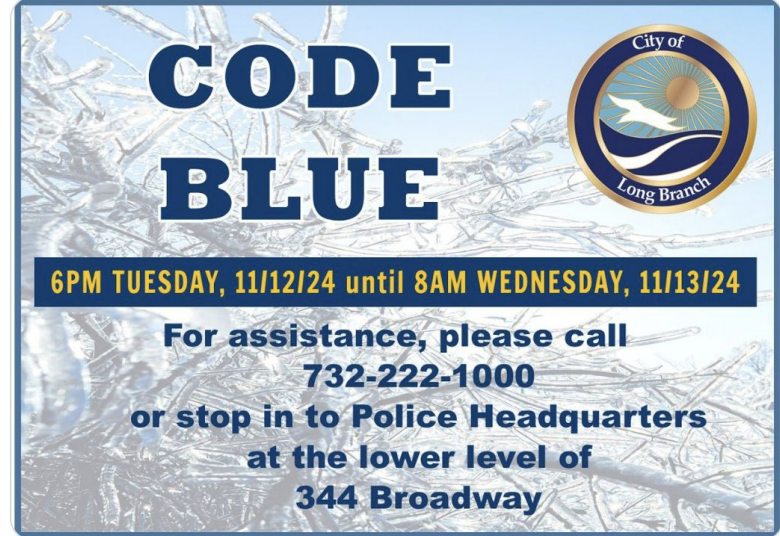


Long Branch Police @LongBranchPD · Nov 12

Code Blue

A Code Blue Alert has been issued for Monmouth County from 6PM 11/12/24 to 8AM 11/13/24

For assistance, please call 732-222-1000 or stop in to Police Headquarters at the lower level of 344 Broadway



60



The CAP Formula

(and proof points)

CARE

Begin your statement with empathy.
Show you care for the people affected by the crisis.

ACTION

Tell people exactly **what you are doing** to fix the problem.
Rather than broad statements such as 'we're doing everything possible', look for specific actions you are taking.

PERSPECTIVE

Finish by putting the issue into some **perspective or context**. I.e. Address how rare the incident is, that you have a good record, that the issue is brand new and couldn't be predicted or is only affecting a small area.

Prepare, listen and adapt

Prepare messaging ahead of time

Listen to responses and adapt messages

Adapt content for different channels

Zack Fradella @ZackFradellaWx

Here is how things will progress tonight into Tue morning. Expect the threat for a few severe storms to pick up over western LA soon. The threat transitions overnight. Everything slows down early Tue morning over New Orleans/coastal MS which is where flooding is possible. #lawx

THIS EVENING



BROKEN LINE OF ST TO SEVERE STORM MOVES INTO WEST LA

Robin Davis @rockinrobin152 · 20h
@ZackFradellaWx how many inches can we expect in Metairie?

1 1 105

Zack Fradella @ZackFradellaWx · 20h
4-6"

2 1 89

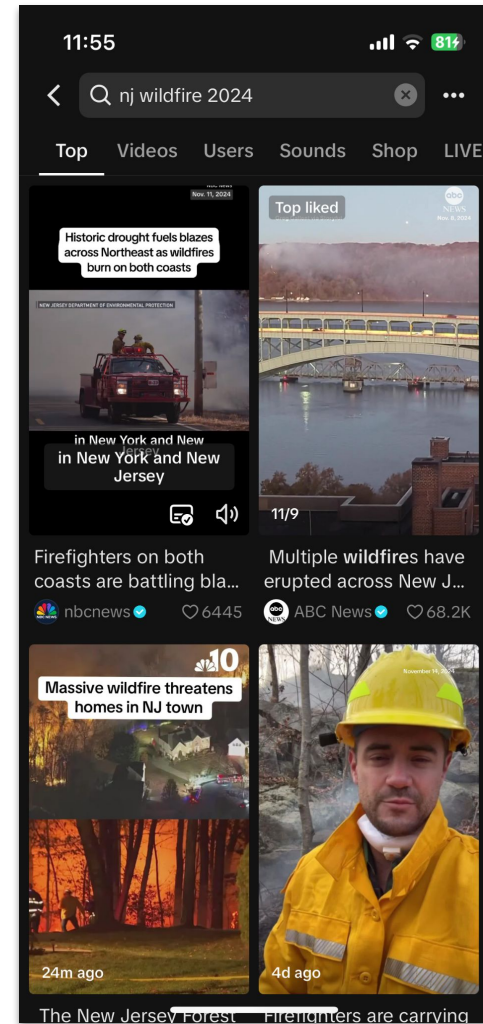
Robin Davis @rockinrobin152 · 20h
Thanks

23

4:00 PM · Nov 18, 2024 · 3,205 Views

Know where your audiences are

- They're not all on one channel
- They're not all following you!
- They are likely using social media as a search engine





Creating impactful content

Why use photo and video?



Quickly explain topics

Spreads further

Echo written messages

Easy to share on

Humanize the response

Grabs attention

How to use graphics

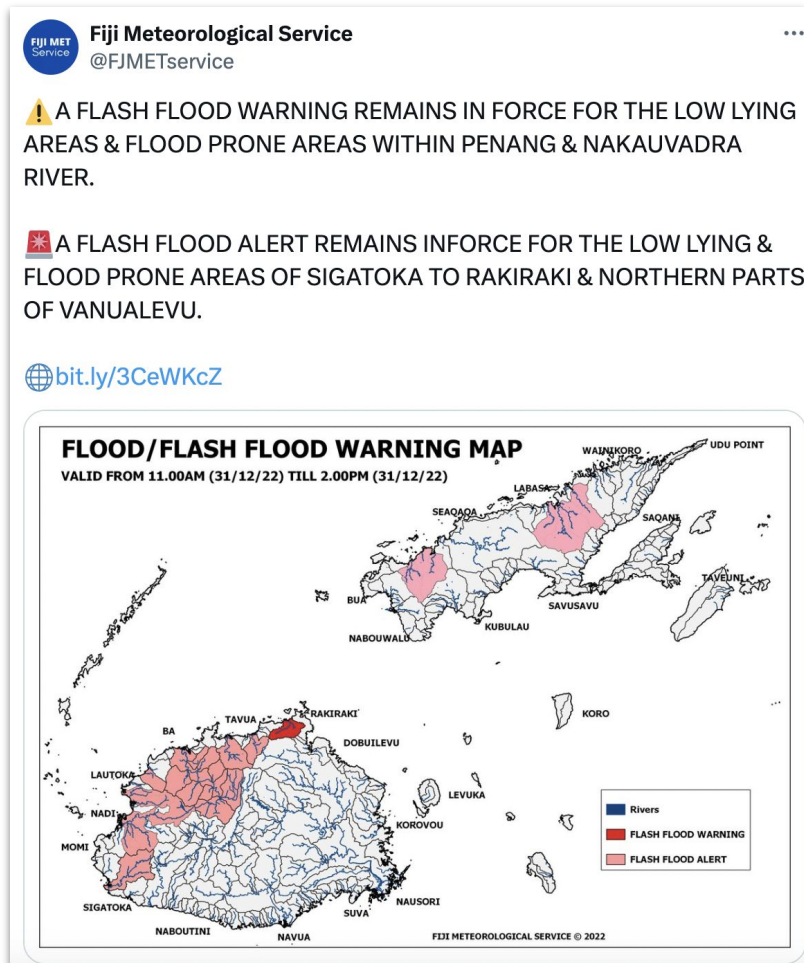
Show a map of an affected area

Provide clear instructions

Share a statement

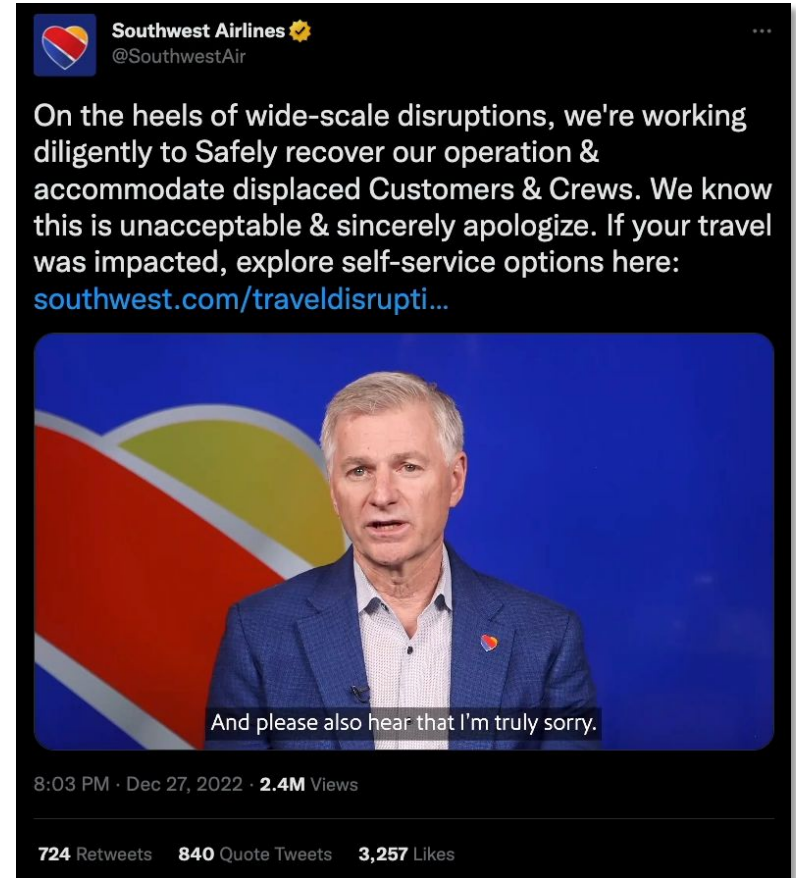
List contact information

Highlight alert levels



Accessibility

- Add alt text to images
- Clear captions for videos
- Don't rely on colour to convey meaning
- Avoid flashing lights in videos, or add a warning if present



Prompt engagement is key



Own the situation

- Show awareness and capability
- Establish your channel as an information source



Shut down misinformation

- Fill the information gap
- Rise above the noise

Responding to misinformation

Explain falsehoods or inaccuracies

Avoid debates

Be clear and succinct

Reply to superspreaders

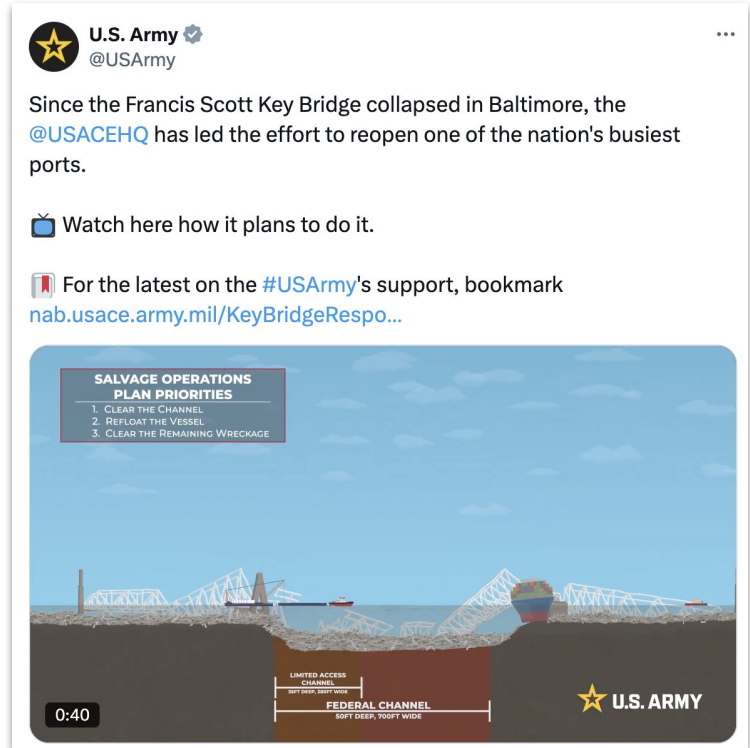
Link to evidence or other sources

Encourage sharing



Leverage third parties

- High levels of trust
- Additional people 'on the ground'
- Expand your reach
- Can provide you with more content



Create FAQs

- Monitor for Qs
- Use popular search terms and key words
- Think AEO and SEO
- Encourage sharing internally and externally
- Use in responses

Crisis90 Phase 2

Phase 2: Feedback

Channels

Reddit, Discord, X, other channels based on engagement and traction, Mayor's own channels

Messaging

Concern: for those affected

Action: response efforts

Perspective: scale of flooding and exactly where affected

Creative Content

Video of impacted area, spokespeople at sites featuring in misinformation posts, shareable by community groups

This is the time to:

- 1 Decide on your crisis plan and escalation process
- 2 Know your audiences and the channels they use
- 3 Check your monitoring keywords and channels
- 4 Connect with potential response partners
- 5 Share knowledge with your team and your audiences

Certificate download from

<https://smsworkshop25.crisis90.com/certificate>

Thank you

Claire Turner

Associate Director

Sophie Pattison

Digital Account Executive



www.socialsimulator.com



claire@socialsimulator.com

sophie@socialsimulator.com



[linkedin.com/company/socialsimulator/](https://www.linkedin.com/company/socialsimulator/)