

The Misinformation Response Gap

When a false narrative becomes an operational problem



ARTEFACT91

Campaign Intelligence + Strategic Communications

artefact91.com

ARTEFACT91 BRIEFING 06

In this briefing

When a false narrative becomes an operational problem

- Executive Brief
- 01 The Narrative Threat Is Not the Claim Alone
- 02 The Artefact91 Narrative Threat Matrix
- 03 Choose the Intervention, Not Just the Message
- 04 Correction Has a Trust Cost If Done Badly
- 05 Build a Misinformation Response Protocol
- 06 Measure Behaviour, Not Volume
- Artefact91 Point of View
- Questions for Leaders
- Sources & Research Base

HOW TO READ IT

A briefing, not a report: the executive brief carries the argument, the numbered sections carry the evidence and the method, and every external figure is attributed to its source.

ARTEFACT91 BRIEFING 06

Executive Brief

Misinformation is not simply a content problem. For institutions, the consequential moment arrives when an inaccurate or misleading claim changes behaviour: employees hesitate, customers cancel, partners seek reassurance, journalists ask questions, regulators take notice, or leadership begins making decisions around a distorted information environment.

The communications challenge is therefore not “How quickly can we correct this?” It is “What kind of information threat is this, what harm can it produce, and what intervention is proportionate?”

Research reinforces the need for discrimination. A 2024 Nature Human Behaviour study across the United States, Poland and Hong Kong found that fact-checking, media-literacy messaging and misinformation coverage reduced belief in false information, but could also increase scepticism toward factual information. A separate 2024 experiment across Germany, Greece, Ireland and Poland found both prebunking and debunking effective, with debunking slightly stronger on some outcomes. The implication is not that correction fails. It is that indiscriminate correction can create second-order effects.

SECTION 01

The Narrative Threat Is Not the Claim Alone

Institutions often assess misinformation by asking whether a statement is true or false. Artefact91 recommends a wider threat assessment.

The relevant variables are:

- REACH — How many relevant people are encountering the claim?
- VELOCITY — How quickly is exposure increasing?
- AUTHORITY — Who is repeating or validating it?
- PLAUSIBILITY — Does it connect to something audiences already suspect?
- CONSEQUENCE — What behaviour could the claim change?
- PERSISTENCE — Is it likely to disappear, recur or become searchable history?
- CORRECTABILITY — Is there clear evidence capable of resolving the central claim?

A false post with large impressions but little behavioural consequence can be less important than a niche claim circulating among a regulator, workforce, investor group or influential professional community.

SECTION 02

The Artefact91 Narrative Threat Matrix

Artefact91 classifies narrative threats using two primary dimensions: evidence of propagation and potential institutional consequence.

LOW PROPAGATION / LOW CONSEQUENCE — Observe. Preserve evidence. Avoid unnecessary amplification.

HIGH PROPAGATION / LOW CONSEQUENCE — Clarify selectively. Strengthen discoverable factual information.

LOW PROPAGATION / HIGH CONSEQUENCE — Intervene directly with affected decision-makers even if public reach is small.

HIGH PROPAGATION / HIGH CONSEQUENCE — Activate a coordinated response across communications, operations, legal, leadership and stakeholder channels.

The matrix prevents “viral” from becoming the only escalation criterion.

SECTION 03

Choose the Intervention, Not Just the Message

The research literature describes multiple interventions: accuracy prompts, rebuttals, friction, inoculation, lateral reading, media-literacy interventions, source labels, warnings and fact checks. A 2024 Nature Human Behaviour review synthesized evidence across 81 scientific papers and identified nine major intervention types.

For an institution, the practical choices can be simplified:

PREBUNK — Prepare audiences for a foreseeable manipulation tactic or misleading narrative before exposure.

DEBUNK — Correct a claim after exposure with clear evidence.

CONTEXTUALIZE — Add missing context when the disputed material is technically accurate but materially misleading.

CONTAIN — Address the audiences or systems through which harm can occur without amplifying the claim publicly.

OUT-EVIDENCE — Publish stronger, retrievable proof that can become the authoritative reference.

DO NOT AMPLIFY — Monitor when intervention would create more reach than the original claim.

SECTION 04

Correction Has a Trust Cost If Done Badly

A correction should not simply repeat the false claim in larger type. It should establish the accurate proposition, provide evidence, explain what is known and unknown, and make the source of verification easy to inspect.

Institutions should also consider messenger credibility. Research across four EU countries found source trust could shape the effectiveness of some debunking interventions. The person with the highest organizational rank is therefore not automatically the most credible corrector.

SECTION 05

Build a Misinformation Response Protocol

Before an issue occurs, establish:

- monitoring and escalation criteria;
- a verified fact repository;
- decision ownership;
- legal and operational verification routes;
- priority stakeholder channels;
- pre-approved correction formats;
- search and AI-answer monitoring;
- criteria for public versus targeted intervention;
- a post-response learning review.

The protocol should distinguish uncertainty from falsehood. An institution damages credibility when it labels legitimate questions or incomplete reporting as misinformation merely because they are inconvenient.

SECTION 06

Measure Behaviour, Not Volume

Do not declare success because the corrective post received impressions.

Track:

- recurrence of the claim;
- movement among priority audiences;
- search and AI-answer changes;
- stakeholder inquiries;
- media uptake of the correction;
- behavioural consequences such as cancellations, complaints, employee questions or partner concern;
- whether the institution's evidence becomes the reference source.

ARTEFACT91 BRIEFING 06

Artefact91 Point of View

The misinformation response gap is the distance between seeing a false narrative and understanding what decision it requires. Closing that gap requires intelligence, not reflex.

Artefact91 Narrative Response Sequence:

ARTEFACT91 WORKING MODEL

DETECT → CLASSIFY → MODEL HARM → SELECT INTERVENTION → RESPOND → VERIFY → ADAPT

ARTEFACT91 BRIEFING 06

Questions for Leaders

What false or misleading narrative could materially change stakeholder behaviour?

What evidence would we need to correct it credibly?

Who would audiences believe?

What would trigger intervention?

When would silence be more responsible than amplification?

ARTEFACT91 BRIEFING 06

Sources & Research Base

Hoes et al. (2024), Nature Human Behaviour — Prominent misinformation interventions reduce misperceptions but increase scepticism.

Bruns et al. (2024), Scientific Reports — Source trust, prebunks and debunks across four EU countries.

Kozyreva et al. (2024), Nature Human Behaviour — Toolbox of individual-level interventions against online misinformation.

Nature Communications (2025) — Psychological inoculation and booster interventions over time.