INTERNATIONAL JOURNAL OF SOCIAL SCIENCES BULLETIN
ISSN: (E) 3007-1917 (P) 3007-1909
Volume 4, Issue 6, 2026

WHEN THE MACHINE SPEAKS SCIENCE: A CRITICAL REVIEW OF GENERATIVE-AI DISINFORMATION, ANTI-SCIENCE COMMUNITIES, AND COMMUNICATIVE RESILIENCE

Umar Nawaz[*1], Areeba Tariq[2], Pawan Masud[3]

*[*1,2,3]MPhil in Journalism and Media Studies Critical Integrative Review In-Charge Department of Media and Communication Studies*

Corresponding Author: *
Umar Nawaz
DOI: https://doi.org/10.5281/zenodo.21352008

Received	Accepted	Published
30 April 2026	15 June 2026	30 June 2026

ABSTRACT

Generative artificial intelligence has made it cheap, kind of easy to produce text, images, audio and video that look real, but really are not. And this kind of capability comes in at a time when public trust in science is already kind of contested, plus organized anti-science groups have gotten pretty good at using digital platforms, with a surprising amount of skill. This review brings together three bodies of scholarship that are usually kept apart: research on AI-generated disinformation, research on anti-science movements and conspiracy communities, and research on communicative resilience. The aim is to understand what happens when synthetic content enters the contest over scientific authority, and what protects publics and institutions when it does. Drawing on a critical integrative method, the review examines literature published mainly between 2021 and 2026 across communication, science communication, and the wider social and behavioural sciences. What's been happening in 2026 across communication, science communication and the wider social and behavioural sciences, there are a few findings that jump out. First, the danger from generative AI isn't mainly about any one fake thing, like a single fabricated report, it's more about the scale and speed, and the personalising part it enables, and also the messy uncertainty it spreads even when no one is fully fooled. Second, a lot of the defenses scholars rely on most—media literacy, fact-checking, psychological inoculation—were mostly designed and tested against human-made, text-based falsehood, so it's still an open problem how well those same tools work when the media is high-quality synthetic. Third, work on citizen level resilience and work on how scientific institutions protect their reputation have been running in almost separate lanes, even though one synthetic attack can test both at once, in the same moment. The review argues that the field needs an integrated, multi-level account of communicative resilience, one that connects the psychology of individual judgement, the design of corrective messages, and the strategies of institutions, and that pays particular attention to AI-generated scientific disinformation and to lower-resilience media environments in Southern Europe. It closes by setting out a research agenda, practical implications for communicators, and policy implications tied to the European Union's emerging regulatory framework.

Keywords: *generative artificial intelligence; disinformation; science communication; anti-science communities; communicative resilience; digital trust; media literacy; platform governance*

1. INTRODUCTION

For most of the last decade, the study of online falsehood seemed to rest on a pretty stable set of ideas. False content was made by people, passed

around through networks, and people believed it for reasons tied to psychology and politics. Researchers looked at how it travels, who is vulnerable, and what could slow it down. Those questions still matter, surely, but now they are inside a bigger shift. Generative artificial intelligence can generate a convincing news report, a fabricated photograph, a cloned voice, or a synthetic video in seconds, and it can do that cheaply as well, plus at scale (Goldstein et al., 2023). The barrier that once separated those with the skill to fabricate from those without it has effectively dissolved. Anyone with a prompt can now manufacture material that looks and sounds real.

This shift matters everywhere, but it matters in a particular way for science. Scientific knowledge relies on trust in this chain of expertise that most people can't actually check themselves, not in any straightforward way. So, when a synthetic video comes along that looks like a respected researcher saying something they never, ever said or when an AI system spits out a believable but fabricated study summary, the normal little cues people use to decide what's credible get kind of flipped, used against them, almost quietly. And it gets worse because anti science movements were already pretty

sharp at working digital platforms even before generative AI arrived. Online spaces built around vaccine refusal climate denial, and pandemic conspiracy have proven they can grow, shift tactics, and still resist moderation (Johnson et al., 2020). At the same time, generative AI basically hands these groups a stronger instrument right when their audience is widest.

Still, it's not just a straight line to doom. A few scholars argue that generative AI might alter the economics of disinformation more than it changes the final outcomes, like lowering the price of making misleading content without automatically boosting how convincing it is, or even how far it travels (Kapoor & Narayanan, 2023). Others point to a growing toolkit of defenses, from media literacy to psychological inoculation, that has shown real promise against misinformation (Roozenbeek et al., 2022). The honest position is that the field does not yet know how the balance between synthetic threat and human resilience will settle, especially in the domain of science. That uncertainty is the motivation for this review.

This article offers a critical integrative review of three literatures that are rarely read together. The first concerns generative AI and the disinformation it enables. The second concerns anti-science communities and the dynamics of distrust. The third concerns communicative resilience, the capacity of individuals and institutions to withstand and recover from manipulation. The argument developed across the review is that these three strands describe a single problem and should be studied as one. When a synthetic artefact attacks scientific authority, it simultaneously tests a citizen's ability to detect a fake, an institution's ability to defend its reputation, and a community's willingness to believe. Understanding resilience therefore requires connecting levels of analysis that current scholarship keeps separate.

The review is also written with a specific scholarly context in mind. It is intended to support doctoral research at the University of Pavia, whose Department of Political and Social Sciences combines strengths in digital methods, public opinion, science communication, and the applied media analysis of the affiliated Osservatorio di Pavia. The Italian and Southern European setting is not incidental. Comparative research also points to the idea that media environments in this part of the world are often more exposed to disinformation than in Northern Europe (Humprecht et al., 2020), which makes the region a useful and strangely under studied place to examine resilience. The review keeps this context in view while engaging with the international literature.

The remainder of the article proceeds as follows. It first describes the review method. It then lays out the theoretical grounding before, you know, looking one by one at the new information ecosystem that generative AI ushers in, the specific headache of AI-generated scientific disinformation, how anti-science communities behave, the part played by

platform algorithms, the current condition of science communication, the notion of communicative resilience, and finally the issues around digital trust and reputational defense. A critical back-and-forth brings these strands together, then the review points to research gaps, suggests a research agenda, and weighs practical as well as policy implications.

2. Background

Now, this whole contemporary worry about disinformation didn't really start with generative AI. It sort of crystallised around the political moments of the mid-2010s, when the spread of false content during elections and referendums sparked a fresh wave of scholarship. One influential take argued that solving what was then called fake news would have to be a coordinated push across fields, since the phenomenon sits right at the intersection of technology, psychology, and politics (Lazer et al., 2018). Around that period, a big study on how stories travel on social media concluded that falsehood moved faster, farther, and deeper than truth, and that false political content spread especially easily (Vosoughi et al., 2018). Those results set two anchors that still guide the field today: first, that the way platforms are built tends to favour novel, emotionally charged material where falsehood often wins the swap, and second, that human attention and judgement are, basically, the real battleground.

Scholars also came up with a kind of shared vocabulary for talking around this issue, more or less. A widely adopted framework for information disorder then breaks things down into misinformation, which is false but not always meant to mislead, disinformation, which is false and also intentionally harmful, and malinformation, which is real information but then weaponised in order to cause damage (Wardle & Derakhshan, 2017). This split matters for the present review because generative AI can end up touching each category in a different way. It drops the "price" of manufacturing deliberate disinformation, it makes the assembly of plausible malinformation easier, and it muddles the boundary between authentic and synthetic that the framework kinda takes for granted.

The science domain has its own distinct history within this larger story. Misinformation about science is not only a matter of bad actors spreading falsehood from outside; it is also entangled with dysfunctions inside the scientific enterprise itself, including hype, selective reporting, and the pressures of publication (West & Bergstrom, 2021). So, in effect the credibility of science can get harmed both from outside pressure and from within weakness, and the two forms can also feed into each other. Once generative AI shows up, it doesn't meet a solid, settled fortress of trust. Instead it arrives in a contested landscape where authority is already negotiated, and sometimes brittle.

It's in that same context that the arrival of capable generative systems should be understood. The technology does not create scientific misinformation, which, as it turns out, long predates it. Instead, it pushes existing weak spots further, and adds new vulnerabilities of its own. At the same time, as later sections will discuss, it can also supply potential instruments for defence. The job of this review is to weigh that balance carefully, rather than just assume the worst, or hope for the best.

3. Review Method

This article follows the logic of a critical integrative review. Unlike a systematic review, which is try to answer one narrow question by really cataloguing studies of just one type, an integrative review i s more like it brings together diverse evidence—empirical studies, theoretical work, and policy analysis—so you can synthesise a scattered field and also spark some new conceptual insight. It fits this project, because the aim is not to pin down one single effect, but rather to tie together literatures that grew apart, and then see where they align, where they clash, and where they just leave open gaps. The critical element signals that the review evaluates and argues rather than merely summarises.

The literature was found using structured searches across major scholarly databases, like Web of Science, Scopus, and Google Scholar,

plus extra targeted searches in prominent journals in communication and science communication. The search terms mixed three concept clusters. One cluster covered generative AI and synthetic media stuff, including generative artificial intelligence, large language models, deepfakes, and synthetic media. Another cluster covered disinformation and its scientific cousin, such as disinformation, misinformation, scientific misinformation, and anti-science. The last cluster related to resilience and defence—so terms like inoculation, prebunking, media literacy, trust, and resilience. On top of that, the process included going through the reference lists of key articles, and using forward citations to track what later papers built on.

The review focused on peer reviewed literature published from 2021 to 2026, which is the window where generative AI shifted from a kind of specialist concern to a more central public issue. Older works were included where they provide the theoretical foundations on which current research builds, such as the original statements of inoculation theory and the information disorder framework. Sources were included if they addressed at least one of the review's three core strands and contributed conceptually or empirically to understanding the intersection of synthetic media, scientific authority, and resilience. Purely technical work on detection algorithms, with no connection to communication or public understanding, was excluded, as were sources that did not meet basic standards of scholarly quality.

Because the review is critical and integrative rather than systematic, the synthesis moves forward kind of analytically. Studies are set side by side across their theoretical assumptions, their methods, their regional settings, and their findings too, with special focus on contradictions and limited scope. The aim throughout is to build an argument about how the field should understand communicative resilience in the age of synthetic science, and to locate the specific gaps that future doctoral research could address. This method has limits, which the review acknowledges: the selection and interpretation of sources involve judgement, and a critical review does not claim the exhaustive completeness of a systematic protocol. What it offers instead is conceptual integration and a clear line of argument.

4. Theoretical Foundations

A review that crosses several literatures needs some theoretical scaffolding that can keep things from falling apart. Instead of hinging everything on just one theory, this article pulls from a bundle of complementary frameworks, each one shining a light on a different slice of the same problem. Put together, they clarify how synthetic content works to persuade, how resistance ends up being built, how the phenomenon ought to be categorised, and how institutions respond once the dust settles.

The first base is the Elaboration Likelihood Model of persuasion (Petty & Cacioppo, 1986). This model separates two pathways through which messages end up changing minds. The central pathway has to do with deliberate thinking about the substance of an argument, while the peripheral pathway leans on surface cues like the apparent authority, or even the attractiveness of a source, depending on the situation. The model is valuable here because it explains why synthetic media can be so potent. A realistic deepfake offers exactly the kind of vivid peripheral cue, the familiar face, the trusted voice, that can bypass careful scrutiny. It also implies that interventions which encourage effortful, central processing should strengthen resistance, a prediction that connects directly to the literature on resilience.

The second foundation is inoculation theory, which comes out of early work on resistance to persuasion (McGuire, 1964) and then, later, got sort of revived and extended in current misinformation research (van der Linden, 2022). The basic suggestion is that when you expose people to a weakened form of a manipulative argument, along with a direct refutation, they become less susceptible to the stronger versions they might meet afterwards. It is kind of like a vaccine for the immune system, except here the protection is more cognitive or

rhetorical. This notion feeds into the fast-growing area of prebunking and it also offers one of the more promising routes toward proactive resilience. For scientific disinformation its relevance is pretty straightforward, because in theory people can be inoculated against the persuasive moves that often show up in anti-science messaging.

The third foundation is the information disorder framework (Wardle & Derakhshan, 2017), which gives the review a conceptual set of words so it stays careful. By teasing apart misinformation, disinformation, and malinformation, and by specifying who is behind the creation of content, what the message is, and who interprets it after, the framework stops the analysis from blending separate issues into one general bucket. That level of separation really matters, especially now that synthetic media makes the line between real and fabricated content, harder to spot, at least in practice.

The fourth foundation concerns the way people respond to content they believe a machine has produced. The MAIN model and the associated notion of the machine heuristic describe the mental shortcuts that come into play when audiences attribute content to a technological source (Sundar, 2008). Sometimes people trust machine output more, like it's objective, and sometimes they distrust it less, like it's inauthentic. As audiences become more aware that content might be AI-generated, and as European law starts requiring labelling of synthetic media, these little heuristics kinda become a main lens for how disclosure will change credibility. This groundwork also ties the psychological side and the regulatory side of the review, pretty tightly, together.

The fifth foundation comes from crisis communication, specifically Situational Crisis Communication Theory (Coombs, 2007). This theory explains how organisations should select response strategies according to how much responsibility audiences attribute to them for a crisis. It governs the review's treatment of reputational defense, allowing the analysis to move beyond the individual citizen to the scientific institutions that must protect their credibility. Crucially, synthetic media strains this theory, because a deepfake can manufacture a crisis around an event that never occurred, forcing an institution to prove a negative before it can manage the fallout.

These five foundations do not compete; they operate at different levels. The Elaboration Likelihood Model and the machine heuristic explain individual processing. Inoculation theory explains how individual resistance can be built in advance. The information disorder framework defines and categorises the phenomenon. Situational Crisis Communication Theory governs the institutional response. The integration of these levels is itself part of the review's argument, since the central claim is that communicative resilience must be understood as a property of individuals, messages, and institutions at once.

5. Generative AI and the New Information Ecosystem

Generative AI changes the information environment in ways that go beyond adding more content to an already crowded space. The most fundamental change concerns production. Where making believable false stuff used to demand human craft, time, and money, generative systems now can make text, pictures, audio, and video whenever, with pretty close to marginal cost (Goldstein et al., 2023). That drop in fabrication cost is arguably the most consequential thing in the new ecosystem, mainly because it removes the real-world limits that used to curb the amount of disinformation.

The evidence on persuasion is sobering. One preregistered experiment found that propaganda produced by a large language model was almost as persuasive as propaganda written by human authors, and that small human editing could basically close the remaining gap (Goldstein et al., 2024). This undercuts a comforting assumption that machine-generated content is somehow thinner or less convincing than human work. The same study noted that the original human propaganda was itself only moderately persuasive, which suggests

that the comparison is not between a powerful machine and a masterful human but between two roughly equivalent sources, one of which can now operate at limitless scale. The implication is that the danger lies less in any single piece of content than in the capacity to produce and target a great deal of it.

This points to a second change: personalisation. Generative systems can tailor messages to individuals or groups, maybe even allowing manipulation to adapt to the specific anxieties and identities of its targets. In that "scale + personalisation" combination there's something that cuts hardest against the old setup, where persuasive messages were mostly one-size-fits-all, and done in a pretty uniform way. The concern is not merely more falsehood but falsehood fitted to each recipient.

Against these alarming possibilities stands an important sceptical position. Some scholars also say that the key binding constraint on disinformation hasn't been the actual supply of content but rather its distribution, plus also the audiences willingness to trust it (Kapoor & Narayanan, 2023). On this reading, generative AI lowers production costs for actors who were already able to produce falsehoods, yet it does not really crack their tougher problem of getting noticed, and then getting people convinced. If that is right, the argument blunts the more apocalyptic forecasts and points to the real battleground as the platforms that spread the content, and the trust ties that shape how it lands for each audience. The review takes this caution seriously, while noting that scale and personalisation may themselves affect distribution and reception in ways that are not yet well measured.

What emerges from this literature is a picture of an information ecosystem in transition, in which the production of convincing content has been democratised and automated, the effects of that change are genuinely contested, and the most important questions are empirical and not yet settled. This unsettled state is precisely why careful research, rather than confident prediction, is needed. The remainder of the review narrows from this general picture to the specific domain of science, where the stakes for public knowledge are especially high.

6. AI-Generated Scientific Disinformation

Within the broad category of AI-generated disinformation, the scientific variant deserves particular attention, and yet it has received less of it than the political and electoral variants that have dominated public debate. This imbalance is itself a finding. The literature on generative AI and falsehood has clustered around elections, foreign influence operations, and, increasingly, health, while the fabrication of scientific content as such, including fake study summaries, synthetic data, fabricated figures, and invented expert statements, remains comparatively under-examined.

The health domain offers the clearest early evidence, and it is instructive. A systematic review of generative AI and health misinformation concluded that these systems can substantially increase the volume, speed, and apparent credibility of false health information, and that ordinary users struggle to distinguish AI-generated from human-authored health falsehood (BMC Public Health, 2025). Health is adjacent to science and shares many of its features, including reliance on expert authority and technical language that lends fabricated content an air of legitimacy. The findings therefore serve as a warning for the scientific domain more broadly: if people cannot reliably tell synthetic health misinformation from the real thing, there is little reason to expect they will fare better with fabricated scientific claims.

A scoping review of generative AI and disinformation across domains helps to map the terrain (López-Borrull & Lopezosa, 2025). It identifies scientific disinformation as one of several distinct thematic areas and emphasises the dual role of the technology, which can both generate falsehood and assist in detecting it. This dual character is a recurring theme in the literature and an important corrective to one-sided alarm. The same systems that fabricate can, in principle, be turned to verification, fact-checking, and the rapid identification of

synthetic content. Whether the offensive or defensive applications will prove more powerful is, again, an open empirical question.

A further strand concerns the changing nature of literacy in this environment. It has been argued that traditional media literacy, focused on evaluating sources and claims, is necessary but no longer sufficient, and that a distinct AI literacy is required to understand and judge machine-generated content (Chu-Ke & Dong, 2024). This argument has particular force in the scientific domain, where the cues people rely on, the presence of citations, technical vocabulary, the appearance of expertise, are exactly the cues that generative systems can now counterfeit. If the markers of credibility can be fabricated, then resilience may depend on a more fundamental understanding of how synthetic content is produced, which is the province of AI literacy rather than media literacy alone.

Comparative work is beginning to appear. A cross-national study of how fact-checking organisations in Brazil, Germany, and the United Kingdom are responding to AI-generated misinformation suggests that practices are still emerging and vary by national context (Cazzamatta & Sarisakaloglu, 2025). This points to the importance of setting: the same synthetic threat may land differently in different media systems, depending on the strength of fact-checking institutions, the level of public trust, and the structure of the information environment. The relative scarcity of this kind of comparative work, especially for Southern Europe, is one of the clearest gaps the review spots. Put all that together, the evidence suggests that AI-generated scientific disinformation is a real and expanding worry, but it is still under- studied as a separate thing. In other words, the domain really needs targeted empirical focus, not just assumptions lifted from nearby fields that kind of overlap.

7. Anti-Science Communities in Digital Spaces

Synthetic content does not enter an empty field; it enters a contest over scientific authority in which organised communities are already active. Understanding these communities is therefore essential to understanding the threat. Research on anti-science attitudes has, moved beyond the older idea that rejecting science happens mainly because people lack knowledge. A more nuanced account says there are several different bases for anti-science belief, like the perceived lack of credibility in the source, friction with a person's social identity, conflict with their established beliefs and values, and even a mismatch between how a message is delivered, and how someone prefers to think (Philipp-Muller et al., 2022). This approach matters because it means that "just give them better facts" will often do nothing, or sometimes even make things worse, when the real reasons for rejection are rooted in identity and values, rather than simple ignorance.

The social structure of these communities is equally important. A network analysis of the online contest between pro-and anti-vaccination views found that anti-vaccination clusters, though smaller than their pro-vaccination counterparts, were more numerous and more effectively entangled with the large body of undecided users (Johnson et al., 2020). The analysis projected that, without any intervention, anti-vaccination views could start dominating the online conversation in the next decade. That reframes everything. The threat is not one solid block of committed believers but more like a shifting and adaptive ecosystem, it is good at working with the persuadable middle. Synthetic content, tailored and produced at scale, is exactly the kind of instrument that could make that engagement more effective.

The adaptive character of these communities is underscored by research on the broader online hate ecology, which found that hostile communities survive moderation by reorganising across platforms, healing their networks after disruption much as a resilient organism does (Johnson et al., 2019). The lesson for science disinformation is cautionary. Interventions that target a single platform may simply displace the problem, and communities organised around the rejection of scientific

consensus may prove similarly difficult to disrupt. Resilience, in other words, is a property not only of the defenders of science but also of its organised opponents, and any realistic strategy must reckon with that fact.

This body of work carries a clear implication for the study of generative AI. If anti-science communities are adaptive, identity-driven, and skilled at engaging the undecided, then the marginal effect of synthetic content may be greatest not among committed sceptics, whose minds are already made up, but in the contested space where opinion is still forming. This suggests that research should focus less on whether deepfakes can convert dedicated opponents of science and more on how synthetic content shapes the judgements of the wavering majority. It is a reframing that the current literature gestures toward but has not yet pursued systematically, and it represents a productive direction for future work.

8. Platform Algorithms and the Amplification of False Scientific Information

If communities supply the social energy behind anti-science content, platform algorithms shape how far and how fast it travels. The popular understanding holds that recommendation systems trap users in echo chambers and filter bubbles, mechanically amplifying whatever provokes engagement, including falsehood. The scholarly picture is more complicated and more interesting, and a careful review must resist the simpler story.

A pretty landmark series of studies, done together with Meta during the 2020 United States election, gives you some of the most rigorous evidence around, and the conclusions are genuinely surprising. In one study they replaced the algorithmic feed with a straightforward chronological one, and that noticeably changed what people were shown, exposing them to more politics and also more stuff from untrustworthy sources, but at the same time it did not really show a measurable shift in political attitudes, knowledge, or polarisation across that three month stretch (Guess, Malhotra, Pan, Barberá, et al., 2023). A companion study found that suppressing reshared content reduced the amount of political news and untrustworthy material in feeds and lowered users' news knowledge, but again produced no detectable effect on polarisation or individual political attitudes (Guess et al., 2023). These results trouble the assumption that algorithmic exposure straightforwardly determines belief.

Two cautions are necessary in interpreting this evidence. First, the studies measured effects over a period of months during a single election, and the absence of a detectable attitudinal effect in that window does not prove the absence of longer-term or cumulative effects. The authors and subsequent commentators have noted that the interventions occurred within an already-saturated information environment, which may limit how much any single change could move entrenched opinions. Second, and crucially for this review, the studies concerned political attitudes rather than scientific beliefs or the specific dynamics of synthetic content. Whether the same null results would hold for AI-generated scientific disinformation, which may operate through different psychological channels, is unknown. The evidence is a valuable corrective to algorithmic determinism, but it should not be over-extended to a domain it did not examine.

What the platform literature establishes, then, is a nuanced position. Algorithms clearly shape exposure, determining what content reaches whom, but the translation of exposure into belief is neither automatic nor simple, and may depend heavily on the content, the context, and the audience. For the study of scientific disinformation, this implies that attention should be paid not only to whether platforms amplify synthetic content but to the conditions under which amplified content actually changes minds. It also implies that interventions focused solely on algorithmic exposure, without attention to the psychological and social processes that govern reception, may disappoint. This is a more demanding research agenda than the echo-chamber story suggests, but it is the one the evidence supports.

9. Science Communication in the Age of Generative AI

The fields surveyed so far describe a threat and its mechanisms. Science communication is the practice charged with responding. Understanding how it is adapting, or failing to adapt, to generative AI is therefore central to any account of resilience. Here the review encounters a genuine thinness in the literature, which is itself worth stating plainly: dedicated, peer-reviewed scholarship on how generative AI is reshaping science communication as a practice remains scarce, even as the broader literatures on AI and on misinformation expand rapidly. This gap is an opportunity for doctoral research.

What can be said draws on the recognition that science communication has been undergoing a longer transformation, of which generative AI is the latest phase. The credibility of science is shaped not only by external attack but by the internal practices of the scientific community, including the tendency toward hype and the distortions introduced by publication pressures (West & Bergstrom, 2021). Generative AI interacts with these existing dynamics in complex ways. So, it can be used to make up fabricated "scientific" content that sounds like it's legitimate, and that's one route. But it can also be used by scientists, and communicators as a kind of translation tool to explain, reframe, and circulate real findings more effectively. In other words, the technology isn't automatically corrosive, and it isn't automatically helpful either. Its impact depends on the way it gets used, and also on how it is governed.

The Italian and Southern European context offers a useful vantage point on these questions, and connects the international literature to a specific setting. There's also research on Italian attitudes from the early Covid 19 emergency, and it showed something kind of double sided. People said they had high trust in scientific institutions, while still holding beliefs that went against the scientific consensus, which kind of warns you off from any simple story where science just "wins" against common sense (Anzivino et al., 2020). A companion analysis then dug into how science, expertise, and everyday common-sense kind of negotiated the meaning of the pandemic—traditional media set the official agenda, while social media directed emotion and counter-narratives (Affuso et al., 2020). These works matter exactly because they don't fall into easy conclusions, they show trust and scepticism can live inside the same public, and that science communication is a contested and emotionally loaded process, not a clean, one-way transfer of facts.

This points toward a particular understanding of science communication's task in the age of synthetic media. The challenge is not only to correct specific falsehoods but to sustain the conditions under which scientific authority can be recognised and trusted at all. Narrative and storytelling, long recognised as central to effective science communication, take on new importance when the factual cues of credibility can be counterfeited, because a coherent and trustworthy account may do work that isolated facts cannot (Ceravolo & Rostan, 2019). The review suggests that science communication research needs to move quickly to understand how generative AI is changing both the threats to scientific authority and the tools available to defend it, and that this is among the most pressing and least developed areas in the entire field.

10. Communicative Resilience

Resilience has become a central organising idea in the study of disinformation, and it is the concept around which this review's argument turns. At the level of whole societies, a influential framework proposed that resilience to online disinformation depends on structural conditions, including the strength of public-service media, the level of trust in news, and the degree of political polarisation (Humprecht et al., 2020). Applying this framework across Western democracies, the authors grouped countries into clusters, with the Northern European nations tending to show higher resilience and the Southern European countries, including Italy, tending to show lower resilience. This structural account is important because it

locates resilience not only in individual minds but in the institutional architecture of media systems, and because it identifies the Southern European setting as a site of particular vulnerability and therefore particular research interest.

At the level of individuals, the most active and promising line of work concerns psychological inoculation, or prebunking. A major set of studies has shown, kind of clearly, that short videos laying out techniques of manipulation could help people recognise those techniques better, and also that this whole approach can be rolled out at scale via online advertising, reaching huge audiences in a real-world field setting (Roozenbeek et al., 2022). The big strength, in my view is that it is proactive, it builds resistance before any exposure rather than trying to correct falsehoods after the fact. It also goes after transferable techniques of manipulation instead of targeting one single, particular claim, which may help it generalise across topics, even when the surface details differ. Reviews in the area have already pulled together a substantial chunk of supporting evidence, and they've begun to map the situations where inoculation tends to work best (Traberg et al., 2022).

A related line of work looks at smaller, more subtle interventions that gently nudge people's attention toward accuracy. Research has shown that something as simple as prompting individuals to think about whether a headline is accurate can improve the quality of the content they later choose to share, with that effect even being seen in a field experiment on a live platform (Pennycook et al., 2021). These accuracy nudges are attractive because they are light-touch and scalable, working with the grain of people's existing desire to share true rather than false content. Alongside them, media literacy interventions have shown that even brief guidance can improve people's ability to distinguish reliable from unreliable news, though with effects that vary across populations and that may fade over time.

Yet the review must register a critical limitation that runs through this entire literature. The defenses in which scholars place most confidence, inoculation, accuracy nudges, and media literacy, were largely developed and tested against human-made, predominantly text-based falsehood. Their effectiveness against high-quality synthetic media, and against AI-generated scientific content specifically, has not been established. There is a real risk that interventions optimised for the previous information environment will prove less potent in the new one, where the cues people are taught to scrutinise can themselves be fabricated. Whether prebunking can be adapted to inoculate people against the techniques of synthetic manipulation, and whether accuracy nudges retain their force when the content in question is a flawless deepfake, are open and pressing questions. The concept of communicative resilience, in short, is well developed in theory but incompletely tested against the threat that now defines the field. Closing that gap is among the most important tasks for future research, and it sits at the centre of the doctoral project this review is intended to support.

11. Digital Trust and Reputation Defence

Resilience and trust are intimately connected, because the ultimate target of much disinformation is not belief in a particular falsehood but confidence in the institutions and sources that allow people to know anything at all. The deepfake literature shows this dynamic with almost unsettling clarity. For example, an experiment using synthetic political video found that people were more likely to feel uncertain rather than thinking they were being actively deceived, yet that uncertainty in itself went on to reduce their trust in news on social media (Vaccari & Chadwick, 2020). So, the damage, was kind of done not because the deception succeeded, but because a broader kind of doubt spread around, one that corrodes trust even when no single, specific lie actually takes hold. This is a profound finding for the study of scientific authority, because it suggests that synthetic media can weaken trust in science without convincing anyone of any particular

false claim, simply by making people unsure of what to believe.

Evidence on detection compounds the concern. A series of preregistered experiments on human ability to identify political deepfakes found that audio-visual cues helped people discern fabricated from genuine content more accurately than transcripts alone, but also that audio produced by state-of-the-art synthesis was especially difficult to detect (Groh et al., 2024). The broader pattern across this literature is that human detection is unreliable and that people tend to overestimate their own ability to spot fakes. If individuals cannot reliably identify synthetic content and wrongly believe that they can, then resilience cannot rest on detection alone, and must instead depend on the kinds of structural and pre-emptive measures discussed earlier.

This is where the question of reputational defense enters, and where the review identifies one of its sharpest gaps. The scientific institutions, universities, journals, and expert bodies whose authority synthetic media threatens are, in effect, organisations facing a novel kind of reputational crisis. Crisis communication theory offers tools for understanding institutional response, explaining how the attribution of responsibility should shape the strategy an organisation adopts (Coombs, 2007). But this theory assumes a real triggering event, and a synthetic attack inverts that assumption by manufacturing a crisis around something that never happened. An institution confronted with a fabricated video of one of its scientists must first establish that the event is false before it can manage its consequences, a task for which existing crisis theory is poorly prepared.

The most encouraging finding in this area is also the one that most directly connects the individual and institutional levels. Experimental evidence indicates that journalistic fact-checks can substantially reduce, and sometimes eliminate, the reputational damage that fabricated scandal videos inflict on an innocent target (Dan, 2025). This matters because it suggests that the corrective which protects an individual's judgement is also the corrective which protects an institution's reputation. The same fact-check works at both levels at once, which is precisely the kind of connection that an integrated account of resilience should seek. Yet research on citizen-level resilience and research on institutional reputation continue to develop in near-isolation, and the literature has not yet built the bridge that this finding implies. Constructing that bridge, theoretically and empirically, is a central opportunity for new research.

12. Critical Discussion

Drawing these literatures together makes it possible to state the review's central argument plainly. The threat posed by generative AI to scientific authority is real but is frequently mischaracterised. It is mischaracterised when it is reduced to the danger of any single convincing fake, because the more profound risk lies in the scale and personalisation of synthetic production and, above all, in the diffuse uncertainty that synthetic media spreads through the information environment. The deepfake research is consistent on this point: the corrosion of trust occurs even without successful deception (Vaccari & Chadwick, 2020). A field that focuses only on whether people can be fooled by particular fakes will miss the larger damage to the conditions of knowing.

The review also exposes a set of contradictions and tensions across the literatures that a synthesis must confront rather than smooth over. There is a bit of a tension here between the worrying evidence about how persuasive AI-generated content can be (Goldstein et al., 2024) and the more sceptical line of thought that the real bottlenecks are distribution, plus trust not production (Kapoor & Narayanan, 2023). At the same time, there is also a tension between the usual echo-chamber story and the more careful platform studies that keep finding that exposure doesn't, really and truly, map onto attitudes in a straight forward way (Guess, Malhotra, Pan, Barberá, et al., 2023). And then, there's the tension between the optimism you

see in the resilience literature and that kind of uncomfortable reality that its preferred interventions still haven't been properly tried against synthetic media. These tensions aren't problems that should be hidden. Instead they point to the active border of the area, and they basically tell you where new research should go first.

A further critical observation concerns method and context. Much of the strongest evidence in this field comes from the United States and Northern Europe, and much of it concerns political rather than scientific content. The platform studies examined an American election; the inoculation studies were conducted largely in English-speaking and Northern European settings; the deepfake experiments used political stimuli. This concentration limits what can be generalised to the scientific domain and to the lower-resilience media environments of Southern Europe, where trust dynamics, media systems, and political cultures differ. The Italian research discussed earlier shows that trust and scepticism can coexist in complex ways within a single public (Anzivino et al., 2020), which warns against importing conclusions wholesale from other contexts. The relative absence of work on scientific disinformation in Southern European settings is not a minor omission but a substantive gap with consequences for how resilience should be understood and built.

Finally, the review's integrative lens reveals the most important structural weakness in the existing scholarship: the separation of levels. Research on individual cognition, research on community dynamics, research on platform effects, and research on institutional reputation each proceed largely on their own terms, with their own theories, methods, and journals. Yet a single synthetic attack on science operates across all these levels simultaneously, testing the citizen's judgement, the community's beliefs, the platform's amplification, and the institution's reputation at once. The finding that a single fact-check can protect both individual and institution (Dan, 2025) is a glimpse of the connections that an integrated account would reveal. The field's fragmentation isn't only some academic annoyance either; it stops the building of resilience strategies that actually work across multiple levels, which is, in practice, what the threat demands. The rest of this review takes those observations, and turns them into a more concrete agenda.

13. Research Gaps

The critical discussion points to several specific gaps that together define a research space. The first is the scientific disinformation gap. The literature on AI-generated falsehood has concentrated on political and, more recently, health content, leaving the fabrication of scientific material, including fake study summaries, synthetic figures, and invented expert statements, comparatively unexamined as a distinct phenomenon. Given that scientific authority rests on exactly the cues that generative systems can counterfeit, this is a consequential omission.

The second is the synthetic-media testing gap. As for the interventions that resilience work tends to lean on—inoculation, accuracy nudges, and media literacy—most of them were checked mainly using human-made, text-based misinformation. But whether they keep working when you add high quality deepfakes and AI-generated scientific content is not clear. So in other words, the current confidence in these defenses may lean on a premise that hasn't been stress-tested yet.

The third is the integration gap. Research on individual resilience and research on institutional reputation proceed in isolation, despite evidence that the same corrective measures can protect both. No established framework connects the individual, message, community, and institutional levels into a coherent account of communicative resilience, even though synthetic attacks operate across all of them.

The fourth is the contextual gap. The strongest evidence derives from the United States and Northern Europe and concerns political content, while the lower-resilience media environments of Southern Europe, including

Italy, remain under-studied. Because resilience depends partly on the structure of media systems and the dynamics of trust, findings from high-resilience contexts cannot be assumed to transfer.

The fifth is the literacy gap. The proposition that AI literacy is a distinct and necessary competence, separate from traditional media literacy, is plausible and increasingly argued, but it has not been firmly established empirically. Whether AI literacy adds explanatory power beyond media literacy in predicting resilience to synthetic content is an open question with direct practical consequences for how interventions should be designed.

14. Future Research Agenda

These gaps translate into a clear agenda. A first priority is set for dedicated empirical research on AI generated scientific disinformation as, like, its own phenomenon. The idea is to look at how synthetic scientific content gets made, how it moves around, and what kinds of reactions audiences show. This kind of work would not just treat scientific disinformation as some "special case" of political misinformation or health misinformation, but instead as a distinct object with its own dynamics, tied to the particular ways scientific authority is set up, and then recognized.

A second priority is, more or less, the experimental testing of resilience interventions against synthetic media, specifically. Studies should line up and compare how inoculation, accuracy nudges, and literacy interventions end up performing when the target content is a high quality deepfake or an AI generated scientific claim, rather than an ordinary
text-based falsehood. That would help show whether existing defenses actually transfer into this new environment, or if they need redesign, and it would test directly whether AI literacy does better than media literacy for building resilience to synthetic content.

A third priority is the development and testing of an integrated, multi-level model of communicative resilience. Such a model would connect the psychology of individual judgement, the design of corrective messages, the dynamics of communities, and the strategies of institutions, and would be tested with research designs capable of capturing effects across these levels. The finding that a single fact-check can protect both individual and institution offers a starting point for this integration, which could be extended through studies that measure citizen resilience and institutional reputational outcomes in tandem.

A fourth priority is comparative and context-sensitive research, with particular attention to Southern Europe. Studies situated in Italian and comparable settings would test whether interventions and dynamics established elsewhere hold in lower-resilience environments, and would draw on the applied media-monitoring capacity of institutions such as the Osservatorio di Pavia to ground the work in real circulation data. This agenda is methodologically ambitious, calling for a combination of experimental, computational, and qualitative approaches, but it is the agenda the threat demands, and it maps closely onto the strengths of a department that combines digital methods, public opinion research, and science communication.

15. Practical Implications

Although this is a review of scholarship, its findings carry practical lessons for those who communicate science and defend its institutions. The first thing is, defense should be proactive rather than just reactive, plain and simple. The evidence that inoculation can build resistance before exposure, and that fact checks can limit reputational harm, indicates that scientific institutions should ready audiences, and response strategies in advance, instead of waiting to correct falsehoods after they've already spread (Roozenbeek et al., 2022; Dan, 2025). Like, pre-emptive talk that explains the methods of manipulation may work better than the after the fact rebuttal, because by then it's harder.

The second lesson is that correction still matters, even when detection is unreliable or messy. The result that fact-checks can reduce, and

sometimes even remove, the reputational damage caused by fabricated content is genuinely encouraging. It also implies that swift, credible correction stays useful, especially for institutions defending against synthetic attacks (Dan, 2025). The main practical hurdle is timing: making correction fast enough, and visible enough, to help before uncertainty really locks in.

The third lesson concerns the limits of literacy as a sole solution. Because people overestimate their ability to detect fakes and because the cues of credibility can be counterfeited, communicators should be cautious about relying on audience literacy alone to carry the burden of defense. Literacy efforts are valuable but should be combined with structural measures, institutional preparedness, and clear labelling of synthetic content. The overall practical message is that resilience is built through a layered combination of proactive inoculation, rapid correction, structural support, and realistic expectations about what any single measure can achieve.

16. Policy Implications

The review's findings are directly relevant to the European Union's emerging regulatory framework, it also makes the policy context sort of unusually concrete. The Artificial Intelligence Act introduces transparency obligations that tell actors to mark AI generated content as artificial and that deepfakes be disclosed, these obligations are set to kick in during 2026 (Regulation (EU) 2024/1689, 2024). The Digital Services Act meanwhile asks big platforms to assess and then reduce systemic risks, including disinformation, and in practice that is the emphasis here. Taken together, these instruments seem to assume that disclosure, labeling, and risk management will protect the information environment , but the evidence reviewed here implies the link between those steps and real resilience is not really firmly established, especially when it comes to synthetic media in the scientific domain.

This gap between regulatory assumption and empirical evidence is itself a policy-relevant finding. Research drawing on the machine heuristic suggests that labelling content as AI-generated could change how people judge its credibility, but the direction and size of that effect are uncertain, and may depend on context and on how the labelling is done (Sundar, 2008). Policymakers implementing the labelling requirements of the Artificial Intelligence Act would benefit from

evidence on whether disclosure actually strengthens resilience or, in some circumstances, inadvertently undermines trust in genuine content. The review thus identifies a direct line between its scholarly agenda and the practical needs of European regulation.

A further policy dimension concerns the coordinated, cross-border nature of the threat and the response. European initiatives addressing foreign information manipulation and interference recognise that synthetic disinformation often crosses national boundaries and targets vulnerable communities, and applied research institutions are increasingly involved in monitoring and countering these campaigns. The involvement of media-monitoring bodies such as the Osservatorio di Pavia in European efforts to counter information manipulation illustrates how academic and applied capacities can be combined in service of resilience. The broad policy implication is that effective response requires evidence-based measures developed in close connection with research, and that the current regulatory moment offers an opportunity to build that connection, provided the necessary evidence is generated in time to inform implementation.

17. Conclusion

Generative artificial intelligence has changed the terms on which scientific authority is contested. By making convincing synthetic content cheap, abundant, and personalised, it hands new capabilities to the anti-science communities that were already adept at exploiting digital platforms, and it threatens to spread a corrosive uncertainty that weakens trust in science even when no particular falsehood

succeeds. This review has argued that meeting this challenge requires reading together three literatures that usually remain apart: the study of AI-generated disinformation, the study of anti-science communities, and the study of communicative resilience. Each illuminates part of the problem, but only their integration reveals its full shape.

The central conclusion is that communicative resilience must be understood as a multi-level property, spanning the judgement of individuals, the design of messages, the dynamics of communities, and the reputational strategies of institutions. The most promising single finding in the literature, that a well-placed fact-check can protect both a citizen's judgement and an institution's reputation at once, points toward the kind of connected account the field needs but has not yet built. At present, the defenses scholars trust most remain largely untested against synthetic media, research levels remain fragmented, and the scientific domain and Southern European contexts remain under-examined. These are not reasons for despair but a map of where work is needed.

The agenda that follows is demanding but coherent. It calls for dedicated study of AI-generated scientific disinformation, for the testing of resilience interventions against synthetic content, for the construction and validation of an integrated multi-level model of resilience, and for comparative research attentive to the particular vulnerabilities of Southern European media environments. Pursued together, these lines of inquiry would move the field from describing a threat to building a defense. As synthetic media becomes an ordinary feature of the information environment, the question is no longer whether machines can speak science convincingly, for they plainly can, but whether publics and institutions can develop the resilience to tell when they should be believed. Answering that question is the work this review has sought to frame.

REFERENCES

Affuso, O., Agodi, M. C., & Ceravolo, F. A. (2020). Scienza, expertise e senso comune: Dimensioni simboliche e sociomateriali della pandemia. Sociologia Italiana – AIS Journal of Sociology, (16), 57–68.

Anzivino, M., Ceravolo, F. A., & Rostan, M. (2020). La rivincita della scienza sul senso comune? Gli orientamenti di fiducia degli italiani all'inizio dell'emergenza Covid-19. Sociologia Italiana – AIS Journal of Sociology, (16), 121–139.

Cazzamatta, R., & Sarisakaloglu, A. (2025). AI-generated misinformation: A case study on emerging trends in fact-checking practices across Brazil, Germany, and the United Kingdom. Emerging Media. https://doi.org/10.1177/27523543251344971

Ceravolo, F. A., & Rostan, M. (2019). Introduzione. Raccontare la scienza: Opportunità e rischi nella società dell'informazione. Cambio. Rivista sulle Trasformazioni Sociali, 9(18). https://oaj.fupress.net/index.php/cambio/article/view/1370

Chu-Ke, C., & Dong, Y. (2024). Misinformation and literacies in the era of generative artificial intelligence: A brief overview and a call for future research. Emerging Media, 2(1), 70–85. https://doi.org/10.1177/27523543241240285

Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of situational crisis communication theory. Corporate Reputation Review, 10(3), 163–176. https://doi.org/10.1057/palgrave.crr.1550049

Dan, V. (2025). Deepfakes as a democratic threat: Experimental evidence shows noxious effects that are reducible through journalistic fact checks. The International Journal of Press/Politics, 31(3), 525–550. https://doi.org/10.1177/19401612251317766

Goldstein, J. A., Chao, J., Grossman, S., Stamos, A., & Tomz, M. (2024). How persuasive is AI-generated propaganda? PNAS Nexus, 3(2), pgae034. https://doi.org/10.1093/pnasnexus/pgae034

Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). Generative language models and automated influence operations: Emerging threats and potential mitigations [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2301.04246

Groh, M., Sankaranarayanan, A., Singh, N., Kim, D. Y., Lippman, A., & Picard, R. (2024). Human detection of political speech deepfakes across transcripts, audio, and video. Nature Communications, 15, 7629. https://doi.org/10.1038/s41467-024-51998-z

Guess, A. M., Malhotra, N., Pan, J., Barbera, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., Gonzalez-Bailon, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., ... Tucker, J. A. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? Science, 381(6656), 398–404. https://doi.org/10.1126/science.abp9364

Guess, A. M., Malhotra, N., Pan, J., Barbera, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., Gonzalez-Bailon, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., ... Tucker, J. A. (2023). Reshares on social media amplify political news but do not detectably affect beliefs or opinions. Science, 381(6656), 404–408. https://doi.org/10.1126/science.add8424

Humprecht, E., Esser, F., & Van Aelst, P. (2020). Resilience to online disinformation: A framework for cross-national comparative research. The International Journal of Press/Politics, 25(3), 493–516. https://doi.org/10.1177/1940161219900126

Johnson, N. F., Leahy, R., Restrepo, N. J., Velasquez, N., Zheng, M., Manrique, P., Devkota, P., & Wuchty, S. (2019). Hidden resilience and adaptive dynamics of the global online hate ecology. Nature, 573(7773), 261–265. https://doi.org/10.1038/s41586-019-1494-7

Johnson, N. F., Velasquez, N., Restrepo, N. J., Leahy, R., Gabriel, N., El Oud, S., Zheng, M., Manrique, P., Wuchty, S., & Lupu, Y. (2020). The online competition between pro- and anti-vaccination views. Nature, 582(7811), 230–233. https://doi.org/10.1038/s41586-020-2281-1

Kapoor, S., & Narayanan, A. (2023). How to prepare for the deluge of generative AI on social media. Knight First Amendment Institute, Columbia University. https://knightcolumbia.org/content/how-to-prepare-for-the-deluge-of-generative-ai-on-social-media

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. Science, 359(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lopez-Borrull, A., & Lopezosa, C. (2025). Mapping the impact of generative AI on disinformation: Insights from a scoping review. Publications, 13(3), 33. https://doi.org/10.3390/publications13030033

McGuire, W. J. (1964). Inducing resistance to persuasion: Some contemporary approaches. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 1, pp. 191–229). Academic Press. https://doi.org/10.1016/S0065-2601(08)60052-0

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. Nature, 592(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 19, pp. 123–205). Academic Press. https://doi.org/10.1016/S0065-2601(08)60214-2

Philipp-Muller, A., Lee, S. W. S., & Petty, R. E. (2022). Why are people antiscience, and what can we do about it? Proceedings of the National Academy of Sciences, 119(30), e2120755119. https://doi.org/10.1073/pnas.2120755119

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). (2024). Official Journal of the European Union, L, 2024/1689. http://data.europa.eu/eli/reg/2024/1689/oj

Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. Science Advances, 8(34), eabo6254. https://doi.org/10.1126/sciadv.abo6254

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), Digital media, youth, and credibility (pp. 73–100). MIT Press.

Traberg, C. S., Roozenbeek, J., & van der Linden, S. (2022). Psychological inoculation against misinformation: Current evidence and future directions. The ANNALS of the American Academy of Political and Social Science, 700(1), 136–151. https://doi.org/10.1177/00027162221087936

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Social Media + Society, 6(1). https://doi.org/10.1177/2056305120903408

van der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. Nature Medicine, 28(3), 460–467. https://doi.org/10.1038/s41591-022-01713-6

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making (Council of Europe report DGI(2017)09). Council of Europe.

West, J. D., & Bergstrom, C. T. (2021). Misinformation in and about science. Proceedings of the National Academy of Sciences, 118(15), e1912444117. https://doi.org/10.1073/pnas.1912444117