Countering Disinformation

Strategies for 2026 and Beyond

A Report on Misinformation Inoculation Studies and Practitioner Guidance

Prepared by Dewey Square Group for Innovating for the Public Good (IFPG)

Executive Summary

In 2024, a joint report written by Dewey Square Group's Digital team and the Digital Democracy Institute of the Americas (DDIA) argued that inoculation can be an important aspect of large-scale disinformation countermeasure strategies. The report included academic evidence supporting inoculation; outlined best practices for forewarning, refutation, and accuracy prompts; and urged civil society organizations to use inoculation countermeasures against disinformation.

In recent years, a growing number of organizations have moved from studying inoculation to deploying prebunking campaigns in practice, including large-scale efforts around elections and on social media.[1] A growing body of evidence shows that well-designed inoculation interventions can meaningfully improve resilience in the face of misinformation.[2]

This paper serves as an update on the latest studies, spotlighting best practices as they continue to evolve. This document serves as a practitioner's guide on how to conduct large-scale disinformation countermeasure strategies. It offers actionable guidelines for crafting

inoculation messages, details the necessity of long-term planning, and outlines proven methods to deploy ongoing "booster shots" over time. [3,4,5]

It is written for civil society organizations, advocacy groups, news organizations, and pro-democracy institutions that run political and issue-based advocacy campaigns.

A note on terminology: We use the closely related terms “inoculation” and “prebunking” throughout this report. **Prebunking**, as we use the term, is any intervention designed to combat misinformation before its intended audience encounters it.

Inoculation is a specific form of prebunking that typically involves two elements: warning people in advance that they may encounter misleading information and giving them a weakened example of the misleading argument or technique along with an explanation of how to recognize and resist it. In academic literature, these two elements are known as “forewarning” and “refutational preemption.”

Key Findings:

- **Inoculation can reduce the spread and impact of disinformation when properly executed, and well-designed strategies can scale to reach large populations.** While there is no “silver bullet” for countering disinformation, inoculation practices are a promising and increasingly well-supported tool.

- **Prebunking has shown effectiveness across the political and ideological spectrum**. It works across ideologies, with studies finding benefits among both Republicans and Democrats and even against misinformation from their own political allies.[6,7]

- **Inoculation is scalable, as demonstrated by large-scale deployments and controlled trials:** a Google/Jigsaw campaign reached more than 120 million YouTube users in five European countries before the 2024 elections, while separate controlled experiments involving nearly 20,000 participants across 12 EU countries found that the prebunking videos improved discernment and sharing decisions.[8]

    - A separate UK paid campaign delivered a prebunking video to more than 375,000 Instagram users and measured the effect inside a live feed.[9]

    - Other researchers also tested AI-generated prebunking on thousands of U.S. registered voters,[10] and tested credible-source corrections and prebunking messages in both the United States and Brazil.[11] Civil society organizations trained thousands of volunteers and pushed hundreds of millions of social media impressions.

Prior research findings:

- **Meta-analytic evidence demonstrated that media-literacy and inoculation interventions can reduce misinformation belief and sharing and improve discernment, although effect sizes vary by outcome and study design.**[12]

- **Inoculation has shown positive effects across different types of misinformation, and some studies find that those effects can persist over time. Research also shows that protection can fade without reinforcement, making "booster shots" an important part of longer-term strategies.[13] For example, one study showed how a paid 19-second Instagram video increased the correct identification of emotional manipulation tactics by 21 percentage points and showed an effect still present five months later, at a cost of roughly $825 per 100,000 impressions.[14] Campaigns should therefore be designed with reinforcement and sustained engagement in mind.**[15]

- "Technique-based" inoculation messaging versus "logic-based" inoculation messaging: this report describes and discusses the benefits of both approaches in recent research.

- Accuracy prompts, recurring spurs for critical thinking, are a complementary intervention that can strengthen some inoculation approaches.

- Local and trusted voices are very powerful as messengers of the inoculation message.

- **Generative AI tools can help produce effective content for inoculation when used carefully.** In the Linegar election study, AI-generated prebunking messages, developed using a carefully designed, human-created prompt, were not detectably less effective than versions that received additional human review.[16] Other research demonstrates the dual-use risk: conversational AI can also be highly persuasive, but efforts to make AI more persuasive can reduce the accuracy of its responses.[17]

- The impact of inoculation messaging can get lost inside noisy feeds. More direct forms of inoculation message delivery designed to cut through social media distractions are recommended.

- A summary of two of the largest meta-analyses of inoculation research to date is included in this document. Those results from dozens of studies involving tens of thousands of participants provide the broadest view yet of the evidence on the effectiveness and limitations of inoculation messaging.

Recommendations for 2026 Midterms and Beyond

Despite the short timeline between now and the general election, **there is still time to implement meaningful inoculation efforts. It is not too late.** Candidates, civil society groups, and organizations can still implement the recommendations in this report.

I**noculation against disinformation is not only valid but crucial.**

Any upcoming communications effort must include it as a central component, with earlier and more consistent implementation preferable to a last-minute effort. This applies across digital advertising, influencer marketing, organic digital outreach, and person-to-person messaging.

These findings could not be more important in the last few weeks before the 2026 midterms and in preparation for the 2028 Presidential election cycle.

This report serves as a core resource for upcoming initiatives, providing current data, recent evidence and best practices.

We also stress how new research underscores that effective inoculation demands a multifaceted strategy. While paid digital campaigns deliver scale, campaigns should also include investing in trusted local messengers and genuine cultural adaptation, co-designing strategies *with* communities rather than *for* them.[18,19]

Table of Contents

What Inoculation Is, and How It Works

Classical inoculation has two core components:

Forewarning: alerting people that they may encounter misleading or manipulative information, creating psychological resistance before exposure.

Refutational preemption: exposing people to a weakened version of a misleading argument or manipulation technique and explaining why it is misleading, giving them practice recognizing and resisting similar claims later.

Much contemporary prebunking is technique-based. Rather than warning people about a particular false claim, technique-based prebunking teaches people to recognize recurring manipulation tactics such as emotional language, false dichotomies, scapegoating, conspiracy thinking, or the use of fake experts, so they can identify those techniques when they encounter them in different contexts.

Accuracy prompting is a complementary intervention that can be paired with inoculation. Accuracy prompts briefly redirect people's attention toward whether information is true before they evaluate or share it. Recent research suggests that combining accuracy prompts with some inoculation interventions can improve truth discernment.

Inoculation interventions can be delivered in almost every way imaginable: as short videos, social posts, games, ad campaigns, workshops, messaging from social influencers, or volunteer conversations. **Ideally, audiences encounter inoculation content before the target disinformation takes hold.** Effective delivery also requires reaching the intended audiences through appropriate channels and, where possible, messengers they trust.

Chapter 1: New Large-Scale Studies: 2022 to 2026

Two years ago, the central question was whether inoculation could work outside of controlled studies and isolated anecdotes. **The evidence increasingly shows that it can: if inoculation is delivered well and follows best practices, it can reach people at scale and improve their ability to recognize manipulation**.

Prebunking U.S. and Brazilian Elections

The most rigorous comparison of prebunking against traditional correction practices comes from Carey, Fogarty, Gehrke, Nyhan, and Reifler (2025), who ran three pre-registered experiments: two in the U.S. around the 2022 midterms and one in Brazil after its 2022 presidential election and the pro-Bolsonaro unrest that followed in January 2023.[20] The experiments tested two approaches head-to-head. The first was retrospective correction from credible sources: Republican politicians and judges speaking against their own partisan interest to affirm the legitimacy of the 2020 U.S. election. The second was prebunking: prospective, factual information about how elections are actually secured: ballot chain-of-custody, audit procedures, fraud-prevention safeguards presented in a "reality vs. myth" format before false claims could take hold.

Both approaches immediately raised election confidence and reduced belief in fraud claims across all three studies.

U.S. Test
Among U.S. Republicans, belief that Joe Biden was the rightful 2020 winner rose from 32.5 percent in the control group to 43.8 percent with credible-source corrections and 38.5 percent with prebunking.

Prebunking has also had an impact across elections: information about the 2022 midterms increased confidence in the 2020 presidential result as well, suggesting that factual knowledge about how elections work transfers broadly rather than applying only to the specific contest addressed. Importantly, the effects of prebunking remained measurable three months later, suggesting that the intervention had some staying power.

Another U.S. experiment by this team tested whether explicit forewarning contributed to the effectiveness of prebunking. Researchers randomized whether participants received a classic inoculation-style forewarning, “here are false claims you might hear,” before the election-security information, or received the same factual content without the forewarning. The study found no clear incremental benefit from adding an explicit forewarning under the conditions

tested. This suggests that, at least in this context, effective prebunking did not require an explicit warning about forthcoming misinformation. Straightforward public education about how elections are secured produced similar effects without the additional forewarning.

Brazil Test
In Brazil, where the political environment was even more acutely polarized, prebunking was the more effective of the two interventions across the measured outcomes: election confidence, fraud beliefs, and accuracy of factual knowledge about how elections are conducted.

Prebunking had a significant effect on all nine outcomes and produced measurably larger effects than credible-source corrections across all nine.

Credible-source corrections only partly worked in Brazil, increasing election confidence and reducing perceived seats won by fraud in 2022, but not changing beliefs about the prevalence of election fraud in either the 2022 or the 2026 election or affecting the expectation of seats won by fraud in 2026. In contrast, prebunking measurably moved the needle on all those factors.

The interventions also produced meaningful effects among strong supporters of right-wing former president Jair Bolsonaro, who lost the 2022 election. Among this group, confidence in the 2022 election rose from 20.0 percent in the control condition to 28.5 percent with credible-source corrections and 28.2 percent with prebunking.

One of the Largest Inoculation Campaigns Run to Date: Google/Jigsaw and the EU

One of the largest real-world deployments was the Google/Jigsaw campaign. Ahead of the 2024 European Parliament elections, the paid campaign ran in five European countries and reached more than 120 million YouTube users. Separately, researchers evaluated national-language versions of the videos in 13 preregistered controlled survey experiments involving nearly 20,000 participants across 12 EU countries.[21] Participants in this study were all aged 45 and up, an intentional decision on the part of researchers since previous studies had mainly used younger participants. Researchers wanted to address a gap in the literature and evaluate the intervention in a population that had been portrayed by some as especially susceptible to misinformation.

The videos targeted three techniques: scapegoating, decontextualization, and discrediting. They improved viewers' ability to distinguish real from manipulative content, to name specific techniques, and to make better sharing decisions. Longer videos (about 50 seconds) were more consistently effective than shorter ones, and effects were especially notable among older adults.

Although effective overall across the EU nations studied, this effectiveness varied based on a number of factors including (but not limited to) national GDP, education, and democratic-governance indices.

- **50-second prebunking videos saw the best results with measurable, consistent effects.** All three long videos significantly improved viewers' ability to distinguish manipulative from non-manipulative content across 12 EU countries and 13 surveys.

- **The prebunking videos produced statistically significant and consistent improvement in viewers' sharing decisions, making them somewhat more likely to share reliable content over manipulative content**.

- **Age, digital literacy, general discernment ability, and voting behavior did not meaningfully moderate effectiveness.** The prebunking videos were more effective for those with higher educational attainment and political tolerance but demonstrated positive effects across all demographic groups.

- **Cross-protection exists.** Learning to spot one manipulation tactic (e.g., decontextualization) also improved detection of unrelated tactics (e.g., scapegoating), suggesting prebunking builds a general "resistance schema," not just topic-specific knowledge.

- **Results varied by country.** Some videos performed far better in some countries than others, underscoring the importance of testing and adapting content for local contexts.

- **Older adults benefited from the interventions.** All participants were 45+, and the videos produced positive effects within this older sample, challenging the assumption that older adults are uniformly resistant to misinformation interventions.

- **"Backfire effects" were generally minor.** Researchers found several small, context-dependent adverse effects, including poorer recognition of some non-manipulative content and, for one long video, increased willingness to share manipulative content. However, these effects were generally minor relative to the positive outcomes, and the authors note that some may reflect methodological artifacts rather than durable harms.

Prebunking Inside a Live Feed: Instagram Field Study

Another significant test came in research published in 2026, when Sander van der Linden and colleagues deployed a 19-second prebunking video as a paid Instagram Story campaign that reached 375,597 U.K. users. Researchers then used Instagram's polling feature to evaluate recognition of emotional manipulation among a self-selected sample of 806 respondents. Van

der Linden, a Cambridge social psychologist and prominent researcher on misinformation and inoculation, appears throughout the research discussed in this report.[22]

Users who saw the video correctly identified emotional manipulation in a test headline 59.6 percent of the time, against 38.2 percent for the control group, a 21-percentage-point lift that still held five months later.

They also found that people shown the inoculation video were three times more willing to click on a “find more” link on the topic than those in the control group, which they said, “suggests that the inoculation interventions can alter (digital) behavior.”

The cost was strikingly low: at about $8.25 CPM (per thousand impressions), reaching 100,000 people cost roughly $825. This evaluation was executed using a fraction of typical political ad budgets. This is the closest study in the new literature to how many digital practitioners actually work: create a short video, buy ads on a platform people already use, and measure results at scale and over time.

The study had important limitations: it used a quasi-experimental design with self-selected poll respondents rather than a randomized population sample, tested a single technique with a single yes/no question, and measured whether people could spot manipulation, not whether they ultimately resisted or declined to share it. **But it provides useful real-world evidence that prebunking can be deployed and measured inexpensively inside a digital environment.**

The researchers summarized: “We provide real-world evidence that brief prebunking videos can be scaled inexpensively and deployed as ads to many thousands of social media users in the context of a scroll feed where attention to accuracy may be limited.” They also stressed “the importance of exploring cultural variation in engagement and effectiveness of inoculation campaigns on global platforms such as Instagram, as the type, meaning, and prevalence of misinformation techniques might differ across platforms and cultures.”

Key Findings

- **Inoculation can work at real-world scale.** Large election and social-media studies show that short prebunking interventions can improve people's ability to recognize manipulation and make better judgments about misinformation.
- **Simple factual prebunking can work.** The election studies suggest that useful new information can sometimes produce benefits even without an explicit warning that misinformation is coming.
- **Hard-to-reach audiences are not necessarily immune.** Positive effects have been found among highly skeptical voters and adults 45 and older.
- **Delivery matters.** Paid placements on YouTube and Instagram demonstrate that prebunking can reach very large audiences, but effects vary by content, audience, country and delivery conditions.

Chapter 2: Studies Showing Best Practices

There are several smaller studies offering key “best-practices” guidelines.

AI (Done Carefully) Can Play a Positive Role in Inoculation Content Creation

The most forward-looking deployment was a pre-registered two-wave experiment by Linegar, Sinclair, van der Linden, and Alvarez (2026) testing AI-generated prebunking against five election-fraud narratives (including claims about voter rolls, voting machine failures, and the "blue shift" in ballot counting) with 4,293 U.S. registered voters surveyed in August 2024, before the general election.[23]

The researchers built their framework on a single insight: if you can develop one good inoculation prompt template, an AI model can rapidly generate effective prebunking messages against any new narrative by swapping in the relevant misinformation and counterfactual evidence. Using Anthropic's Claude 3.5 Sonnet paired with election-security fact sheets from CISA (the Cybersecurity and Infrastructure Security Agency), the researchers refined a template on one myth and then used that template to generate inoculations against the remaining four. The authors likened their template approach to mRNA vaccines, a stable structure rapidly adaptable to new narratives, or "inoculation at machine speed."

Key Findings

- **The AI-generated messages significantly reduced belief in election myths, with the protective effect persisting at least one week.** To illustrate the scale of protection: control-group participants who read a misleading article without inoculation became one point more confident in their belief of false statements about voter-roll fraud on a 10-point scale, while **inoculated participants barely moved.**

- **The study found no evidence of a backfire effect.** No negative effects of the inoculation messages were detected.

- **After researchers used substantial human expertise to develop and refine the AI prompt, AI-generated messages were not detectably less effective than versions receiving additional human review.** The rumor-reduction effects also held across partisan lines, working similarly for Democrats, Independents, and Republicans. This suggests AI could substantially accelerate content production once organizations have developed and validated a strong underlying template.

Unlike the Brazil study above, this study found that while Republicans became less likely to believe specific election rumors, the intervention did not meaningfully increase their overall confidence in election administration, a gap that matters for practitioners targeting the most skeptical audiences. **Further research is needed to understand this gap in findings.** In other words, as the authors note, **factual prebunks may be sufficient to counter specific false claims but not deep-seated institutional skepticism**, which may require more intensive or longer-term interventions.

But AI Is a Two-Edged Sword in Persuasion

As AI chatbots have risen in popularity, a related approach to countering misinformation has emerged: persuading people after the fact through AI-to-human chatbot dialogue.[24] A team led by Thomas Costello, Gordon Pennycook, and David Rand showed that a back-and-forth conversation with an AI chatbot reduced a person's belief in a conspiracy theory by about 20 percent on average, with the effect still measurable two months later and holding even among committed believers.[25,26] The researchers reported that this reduction "was consistently observed across a wide range of conspiracy theories, from classic conspiracies involving the assassination of John F. Kennedy, aliens, and the illuminati, to those pertaining to topical events such as COVID-19 and the 2020 US presidential election." The researchers also noted that this experiment did not have a "backfire" effect of dissuading belief in actual, legitimate conspiracies. This is debunking, the mirror image of inoculation on the timeline but using some of the same overall principles.

The authors note that the study demonstrates "the persuasive power of LLMs" and underscores "the potential positive impacts of generative AI when deployed responsibly." Given how persuasive these systems can be, they also emphasize "the pressing importance of minimizing opportunities for this technology to be used irresponsibly."

An important caveat: In June 2026, Science issued an Editorial Expression of Concern regarding reporting inconsistencies and dataset errors in the above research. The notice says the authors report that their corrected pipeline produces results matching the original in direction, statistical significance, and substantive size, although Science was still evaluating them as of this report's publication.[27] Therefore, the study cannot be considered to be settled evidence until the journal resolves the notice.

Accuracy Prompts Boost Inoculation

A separate line of research has examined the efficacy of accuracy prompts, simple reminders to think about accuracy, when used along with inoculation to improve the quality of what people share online.

In 2024, two research teams that had publicly disagreed about whether accuracy prompts work for conservatives agreed to settle the question together. The Pennycook and Rand team at MIT and Cornell had found that the prompts worked across the political spectrum; the Rathje, Roozenbeek and van der Linden team at NYU and Cambridge had argued the effect was weak or absent for conservatives.

Rather than continuing to publish competing findings, the teams designed a single pre-registered study together: an adversarial collaboration, a particularly rigorous approach in which both sides agree on the methods before seeing the results. Pooling 21 experiments and nearly 28,000 participants, they analyzed the data using 70 different statistical methods. **In every single one of their joint experiments and data analyses, accuracy prompts significantly improved sharing discernment among Republicans and conservatives, meaning they shared more true content relative to false content.**

The improvement ranged from **1.7 to 6.1 percentage points** depending on the type of prompt. **Importantly, the gains were driven primarily by a reduction in sharing false news:** among Republicans, false news sharing dropped between 3.3 and 14.6 percent relative to control groups.

The effect was present across ideologies but was somewhat weaker on the political right than on the left. However, the partisan gap depended on how partisanship was measured. When people were sorted by party affiliation or Trump vote, Democrats showed a stronger response. **But when people were sorted by how conservative they described themselves on a scale, the gap largely disappeared.** This suggests that accuracy prompts bump up against partisan identity more than political ideology.

The bottom line, endorsed by both teams: accuracy prompts work across the political spectrum, though they may work somewhat better for Democrats.[28]

Inoculation Can Get Lost Inside Noisy Feeds

The studies above measure inoculation on something like a quiz: watch a video, then rate a set of headlines. An actual feed is nothing like that, and the effects do not always survive the move. Across five pre-registered studies with nearly 4,800 participants, Wang and colleagues (2025)

tested a validated inoculation video inside a simulated scroll feed. When the researchers mixed the inoculation content with ordinary tweets, or with more than one manipulation technique, the effect disappeared.[29]

Two preregistered field studies on X tested whether a validated inoculation video changed users' subsequent real-world posting and sharing behavior. They found no meaningful effect in the predicted direction. The authors caution, however, that platform ad delivery made verified exposure extremely low, limiting the test and illustrating the practical difficulty of translating controlled efficacy into measurable platform behavior.[30] **Inoculation's own leading researchers now say plainly that doing well on a lab quiz is not the same as changing behavior in a live feed.**[31]

The practical lesson is not that inoculation fails, but that passive organic exposure in crowded feeds may be insufficient. Practitioner experience and broader research suggest reaching audiences through intentional channels, such as targeted video placements, direct digital outreach, or trusted messengers, rather than relying on organic timeline reach. For instance, researchers have recommended deploying short prebunking videos as unskippable pre-roll ads on YouTube, helping ensure that viewers are exposed to the content.

The Effects of Inoculation Fade Without Reinforcement

Numerous studies have found that the effects of inoculation fade without reinforcement. Inoculation decays, and the mechanism of decay reveals something important about what lab studies have actually been measuring.[32]

The clearest demonstration comes from Capewell, Maertens, Remshard, and colleagues (2024), who ran two pre-registered longitudinal studies with a total of 1,176 participants. In the first study, participants watched a validated inoculation video. Some then immediately completed a quiz evaluating social media posts for manipulation (the standard posttest used in most inoculation research); others did not take this quiz until 48 hours later.

The results were stark: participants who took the immediate quiz and then were tested again at 48 hours still significantly outperformed the control group, with a moderate effect size. But participants who received no immediate quiz, who simply watched the video and went about their day, as any real-world viewer would, showed zero difference from controls just 48 hours later. The previously observed effect was no longer detectable.

The second study replicated this pattern with a gamified intervention (the *Bad Vaxx* game) over a 10-day delay. Again, participants who played the game and immediately took a quiz retained measurable gains at 10 days; those who played the game without an immediate quiz did not. **The researchers identified the mechanism: the immediate quiz strengthened participants'**

memory of the intervention's lessons. The quiz was not just measuring learning: it was *reinforcing* it. Participants who took the quiz could recall the specific techniques they had learned about significantly better at ten days than those who had not, even though both groups had received the same inoculation. [33]

A larger follow-up by Maertens and colleagues (2025), running five pre-registered experiments with nearly 12,000 participants, mapped the decay curve in more detail. Text-based and video-based inoculation interventions could remain effective for roughly a month when the study design included an immediate posttest. Gamified interventions' effects diminished more rapidly; the effect of a game-based inoculation was no longer significant just nine days later.

Booster interventions administered at the 10-day mark did help: participants who received a booster still showed a medium-sized effect at 30 days.

The driving mechanism was again memory, not motivation. People's self-reported motivation to resist misinformation was unaffected by boosters, but their ability to remember what they had learned was significantly strengthened.[34]

The practical takeaway is that one-time campaigns are not enough. Any deployment that lacks a mechanism for rehearsal, which describes nearly every real-world campaign, should expect rapid decline in retention. Durable protection requires either active engagement during the intervention (quizzes, practice, feedback), periodic booster doses, or both.

The field's own leading researchers now frame this explicitly: the "memory-motivation" model proposed by Maertens and colleagues holds that inoculation works through strengthened memory, and memory decays without reinforcement.

Local Trusted Voices Matter

Trusted local sources make a difference. A Stanford study confirmed what community organizers have long suspected: who delivers the intervention and how people are recruited matters as much as the content itself.[35] Angela Y. Lee, Ryan C. Moore, and Jeffrey T. Hancock evaluated digital media literacy workshops designed by PEN America in partnership with community organizations serving four communities of color in the United States: Black, Latino, Asian American/Pacific Islander, and Native American. The workshops were tailored to each community's specific information environment and experiences with misinformation.

The researchers ran two studies (total N = 370) to compare the efficacy of messenger types. In the first, a quasi-experimental field study, participants were recruited through community outreach by the trusted local organizations. The second study was a randomized controlled trial that recruited Latino participants through a survey company. Both groups improved their

comprehension of digital media literacy concepts after the workshop. **However, only the community-outreach group improved their actual ability to correctly identify true and false news in a behavioral detection task, the skill that matters most in the real world.**

A recent project that harnesses both the power of AI and authentic local voices was undertaken by Onyx Impact, a nonprofit that built a layered information infrastructure specifically for Black communities, which research identifies as disproportionately targeted by coordinated disinformation. The studies reviewed here make clear that effective inoculation requires intentional delivery channels, trusted messengers, and reinforcement mechanisms working together. Onyx Impact's model attempts to address all three simultaneously.

Onyx Impact's Digital Green Book, launched in March 2025 and inspired by the original Green Book that guided Black travelers to safe spaces from 1936 to 1966, delivers an AI-powered platform trained on vetted Black news sources and civil rights history, rather than the open web - a deliberate choice to avoid reproducing stereotypes baked into general-purpose AI tools and models. That sourcing decision directly addresses the trusted-messenger problem: the platform's credibility draws from the same community knowledge networks the Stanford study identifies as essential to applied learning.

Onyx Impact's award-winning Information Integrity Lab, launched in August 2025, extends that logic into human infrastructure, matching Black digital creators with legacy Black newspapers through a partnership with the Black Press, building a resilient information layer before disinformation spreads, and training young people as digital navigators for their communities.

More recently, in June 2026, Onyx Impact launched Aisha, a full AI chat assistant available on iOS and Android that centers Black sources, history, and lived experience. Its prompt system, developed by a Black-led engineering team over more than a year, parallels the Linegar et al. finding that a carefully designed human prompt template enables AI to generate effective responses to new narratives at scale. As a persistent conversational tool, Aisha functions as a reinforcement channel: users return to it repeatedly, addressing the decay dynamic that one-time campaigns cannot. Aisha was not designed or evaluated as a formal inoculation intervention, and its architecture prioritizes user privacy by not using conversations to train its model. Where mainstream AI tools have shown documented bias against Black users, Aisha operates within boundaries designed to reduce bias and harm and cites Black news and vetted sources in its answers.[36]

The implication for practitioners is not that inoculation can only work via trusted local messengers: many successful interventions reviewed in this report were delivered through large-scale digital channels without them. Rather, these findings suggest that, when feasible, trusted community channels and culturally relevant delivery may strengthen an intervention, particularly the ability to translate media-literacy concepts into applied judgment.

Participants recruited through community organizations demonstrated stronger applied learning than those recruited through survey panels, even when both groups understood the material equally well. Community context, trust, and cultural relevance appear to add an effectiveness layer beyond the intervention itself. Onyx Impact's model, combining community-rooted messenger networks, culturally grounded AI, and persistent digital infrastructure, offers one potential framework for organizations seeking to preserve those advantages at scale.

For practitioners, the goal remains building direct audience connections through intentional channels: newsletters, targeted communities, trusted messengers, and platforms audiences already inhabit and trust. For organizations operating at scale, the challenge is how to preserve some of these advantages when genuinely local recruitment is impractical.

Both Technique-Based and Logic-Based Prebunking Can Improve Recognition of Manipulation

In two pre-registered experiments with more than 3,100 participants, Biddlestone and colleagues tested six short animated prebunking videos. Three targeted manipulation tactics commonly used in misinformation: polarization (creating an “us versus them” divide), conspiracy theories (invoking unsupported secret plots), and fake experts (using unqualified or misleading authorities). The other three targeted logical fallacies: whataboutism (deflecting by pointing to others’ wrongdoing), moving the goalposts (changing the standard once it has been met), and the strawman fallacy (distorting an argument to make it easier to attack).[37]

The videos were professionally animated, roughly 60 to 90 seconds long, and followed a consistent structure: define the tactic, warn the viewer, show a humorous example, then demonstrate real-world instances.

The tactic-based videos (Study 1, 1,583 UK participants) all made people better at spotting manipulative content when they saw it. The polarization video had the largest effect: people who watched it rated manipulative polarization posts roughly four-tenths of a point higher on a seven-point scale of perceived manipulativeness than the control group. Videos prebunking conspiracies and fake experts showed smaller but still significant effects.

Crucially, though, only the polarization video improved discernment (in other words, the ability to tell manipulative content apart from legitimate content on the same topic). Videos prebunking conspiracies and fake experts raised people's ability to spot manipulation, but they did not significantly help people distinguish manipulation from straightforward content.

The logic-based videos (Study 2, 1,603 U.S. participants balanced between Republicans and Democrats) performed less consistently. All three improved detection of fallacious content, but again, only one (the strawman fallacy video) improved actual discernment. The “whataboutism”

video showed a similar pattern to the weaker tactic-based videos: better detection of that particular tactic, but no improvement in telling fallacious from non-fallacious content.

The moving of the goalposts video was the outlier, and not in a good way. It made people more suspicious of *everything* related to the topic, including content that contained no fallacy at all. People who watched it rated non-fallacious content as more manipulative than the control group did, the opposite of the desired effect. Rather than teaching people to distinguish bad arguments from good ones, it appeared to teach a generalized distrust. The researchers flagged this as potentially "harmful to the cause of educating viewers" and noted that participants who watched it also reported more counterarguing against the video itself, suggesting the concept may have been harder to absorb or may have primed a reflexive skepticism.

Why did the tactic-based approach outperform the logic-based one? The researchers offered several explanations. Logical fallacies are inherently harder to apply in real-world contexts than in textbook definitions: a “whataboutism” in an actual political argument is harder to spot than one in a logic class. Real-world content often contains multiple overlapping fallacies, making clean identification difficult. And there was far more variability in how well individual test items worked in Study 2, suggesting that logic-based detection depends heavily on the specific example, while manipulation tactics are more consistently recognizable.

The practical implication: most of the videos improved detection of the targeted tactic, but the polarization and strawman videos produced the clearest gains in actual discernment, while the moving-the-goalposts video was the least promising. The researchers recommend deploying the strongest-performing videos as unskippable pre-roll ads and emphasize the need for booster doses as effects decay over time.

Key Findings

- **Don't rely on a one-time intervention.** Inoculation effects can fade quickly; repetition, boosters, quizzes and other opportunities to rehearse what people learned can extend their effects.
- **Design for the environment where people will encounter the message.** Results achieved in controlled experiments do not necessarily survive crowded social-media feeds, making placement and verified exposure important.
- **Technique-based prebunking has particularly strong evidence.** Teaching recognizable manipulation tactics can improve detection and, for some techniques, genuine discernment—but not every tactic or execution works equally well.
- **Combine approaches where appropriate.** Accuracy prompts, culturally relevant delivery and trusted community channels can complement inoculation rather than requiring practitioners to rely on one intervention alone.
- **AI offers scale—with caveats.** AI can rapidly adapt a tested prebunking structure to emerging narratives, but its persuasive power makes human review and safeguards essential.

Chapter 3: New Meta-Studies, or "Studies of Studies" on Inoculation

Finally, we reviewed several wide scale "studies of studies" to see how these findings hold up across the entire field.

A 2026 Study Re-Examines 33 Inoculation Studies

In 2026, a team led by Almog Simchon at Ben-Gurion University, working with prominent misinformation researchers Stephan Lewandowsky and Sander van der Linden, published a large reanalysis in Current Opinion in Psychology. They applied signal detection theory across 33 inoculation experiments with more than 37,000 total participants, using hierarchical Bayesian modeling, a statistical approach that looks for an overall pattern across many studies while also accounting for differences among individual studies and participants.[38]

The analysis found that inoculation interventions consistently improved discrimination (people's ability to correctly identify unreliable content) without increasing response bias. These improvements appeared across gamified, video-based, and hybrid interventions. That is, inoculated people got better at spotting manipulation without becoming more distrustful of legitimate news.

The Simchon meta-analysis provides one of the largest recent reanalysis specifically examining whether inoculation improves genuine discrimination rather than simply increasing skepticism. Its preregistered analysis covered 33 experiments from 13 papers involving 37,025 participants, substantially expanding on an earlier critical reanalysis that examined eight experiments from five papers.

The Largest Recent Synthesis: Dosage Matters

A 2024 meta-analysis by Guanxiong Huang, Wufan Jia, and Wenting Yu at City University of Hong Kong and Hong Kong Polytechnic University pooled 49 experiments with 81,155 total participants; this is one of the most comprehensive quantitative syntheses of media literacy and inoculation interventions to date. The headline finding: these interventions work, and they work across multiple outcome measures. The study measured three potential effects of inoculation interventions separately: whether people believed misinformation, whether they could tell it apart from real news, and whether they would share it. All three improved.[39]

- **The largest effect was on reducing misinformation sharing, followed by improving people's ability to distinguish true from false; the effect on false belief was smaller.**[40]

- **Multi-session programs outperformed single-session interventions.** A one-time prebunking video or game helps, but repeated engagement produces meaningfully stronger results, reinforcing the case for booster doses, much like the medical vaccine analogy that inoculation theory draws on.

- **Cultural variation exists.** Interventions were effective cross-culturally but were more effective in cultures with higher uncertainty avoidance, meaning cultures in which people tend to be less comfortable with ambiguity and uncertainty.

- **College students showed stronger effects than adults** recruited from online crowdsourcing platforms like Amazon Mechanical Turk, likely reflecting differences in cognitive engagement and the artificial nature of paid surveys versus structured educational settings.

Key Findings

- **The broader evidence supports inoculation.** Across large syntheses involving tens of thousands of participants, inoculation and media-literacy interventions improve people's resilience to misinformation.

- **Better discernment does not simply mean greater skepticism.** The Simchon reanalysis found that inoculation improved people's ability to distinguish unreliable from legitimate content without making them broadly distrustful.

- **Dosage matters.** Huang's meta-analysis found stronger effects from multi-session interventions than from single-session approaches.

- **Results vary.** Effects differ by outcome, intervention, audience and cultural context—another reason to avoid treating inoculation as a one-size-fits-all solution.

Chapter 4: Updated Practitioner Guidance

Our new findings build on the classical forewarning-refutation framework, while also considering complementary interventions such as accuracy prompting.

Revisiting the Core Framework: Recommendations

Forewarning should not be abandoned but should instead frame the threat around techniques rather than actors: "You may see content that uses [tactic]" rather than "groups are spreading misleading content." Practitioners should test their language and messaging with the target audience before scaling and consider whether a prebunk without explicit forewarning might perform better in their context.

Refutation should also center tactics rather than formal logic.
The studies outlined above generally found more consistent results from showing people what manipulation looks like in practice, such as fake experts and conspiracy thinking, than from teaching formal logical fallacies.[41]

Practitioners should continue to integrate accuracy prompts. The evidence for accuracy prompts remains strong and now extends across partisan lines.[42] Practitioners should weave them into refutation content as a matter of course. Recent research also suggests that combining inoculation with complementary interventions such as accuracy prompts can be more effective than relying on inoculation alone.[43]

Practitioners should plan for the decay of effects and booster shots from the start.
Because protection fades without reinforcement,[44] practitioners should build sustained sequences, not one-shot pushes, and be aware that multi-session programs outperform single sessions.[45,46] Where possible, they should also include an immediate interactive element such as a short quiz, which research suggests can slow decay.

AI as a Resource for Smaller Organizations

The finding that AI-generated prebunking content was shown to match human-reviewed content in effectiveness in the Linegar election study suggests a meaningful opportunity for resource-constrained organizations.[47] The key is the template technique: develop a stable inoculation structure and use AI to adapt it rapidly to emerging narratives, the "mRNA vaccine" model.

A 2026 framework outlines seven roles generative AI can play, from monitoring narratives to drafting content:

- **Informer:** Retrieves, summarizes, and tailors information.
- **Credibility guard:** Detects, flags and helps identify suspect content.
- **Persuader**: Tailored, dialogue-based generative AI to counter misperceptions.
- **Integrator:** Mediator that synthesizes views into balanced, common-ground summaries.
- **Collaborator:** A copilot in inquiry, organizes sources and scaffolds questions.
- **Teacher:** Personalized tutoring with formative feedback on information evaluation and reasoning.
- **Playmaker:** Maker of adaptive game scenarios to practice verification and prebunking.[48]

Guardrails matter. Review AI output for accuracy before deployment, because the political and reputational stakes of an inaccurate prebunk are high. Treat AI as a drafting tool, not an autonomous pipeline.

Messengers, Delivery, and Reaching Specific Communities

When possible, organizations should invest in local trusted messengers (community leaders, local journalists, faith leaders), volunteer networks like the Truth Brigade, direct community delivery through workshops and in-person events, and local-media partnerships to replace declining platform distribution. Trusted influencers and organizers can get the inoculation messages out in ways that evade noisy and confusing feeds and reach audiences more directly.

Understanding an audience and its culture is important when designing an intervention. The evidence is now strong that general-audience interventions do not automatically transfer across communities, underscoring the value of testing and adapting messages for the people they are intended to reach.

Key Findings

- **Teach tactics, not villains.** Help audiences recognize recurring manipulation techniques rather than simply warning them about particular actors or claims.
- **Build reinforcement in from the beginning.** Plan sustained sequences, practice opportunities and booster messages rather than one-time exposures.
- **Pair inoculation with complementary tools.** Accuracy prompts can strengthen the broader intervention, while quizzes and other interactive elements can reinforce learning.
- **Use AI to increase capacity, not replace judgment.** Stable templates can allow organizations to adapt prebunks rapidly, but human review and accuracy safeguards remain essential.
- **Design with the audience, channel and community in mind.** When feasible, culturally relevant content, trusted messengers and direct audience relationships can strengthen delivery and applied learning.

Chapter 5: What's Next: Building for 2026 and 2028

The 2026 midterms will be the first major U.S. cycle conducted in this new information environment of weakened platform cooperation, evidence-backed but imperfect tools, new AI capabilities, and a fragmented regulatory floor.

For 2026, there is still time to act. Inoculation efforts take time and planning, but the remaining weeks and months give organizations time to put the best practices from this report into action during the final stretch of the midterm campaign.

But even more importantly for 2028, now is the time to build capacity and begin strategizing to incorporate inoculation as a central communication goal.

As generative AI becomes increasingly central to the information environment, organizations will need scalable ways to respond quickly to emerging misinformation. AI can also be part of the defense: with careful human guidance, AI-generated prebunking messages can be as effective as human-reviewed versions, allowing organizations to respond more quickly to emerging misinformation.[49]

Recommended operational priorities:

- **Invest in Organizational Infrastructure:** Recognize that recruiting talent, establishing strategic partnerships, and scaling content pipelines demand long-term preparation.
- **Foster Cross-Sector Alliances:** Collaborate across sectors to aggregate audience reach in an era of fragmented social media distribution.
- **Anchor Messaging in Nonpartisanship:** Protect the broad efficacy of prebunking and inoculation by focusing on democratic integrity rather than partisan outcomes.
- **Develop AI Capabilities and Guardrails:** Proactively integrate and carefully test generative AI models to build mature, reliable operational capacity for AI-enabled inoculation work ahead of the 2028 cycle.

What Organizations Should Do Now

START TODAY Inoculation works best before false narratives become entrenched, making early planning essential.

THINK IN CAMPAIGNS, NOT POSTS. The evidence favors sustained interventions, reinforcement, and repeated engagement over one-off communications.

BUILD FOR THE NEXT CYCLE, NOT JUST THIS ONE. Use 2026 to deploy, test and learn while building the organizational capacity, partnerships and AI-enabled tools needed for 2028.

Conclusion

Inoculation for information integrity works across ideologies, can be deployed inexpensively at scale even inside a live feed, can be accelerated with AI, and can be deployed through channels organizations control directly, providing options that do not rely exclusively on platform distribution.

But there are challenges: inoculation has demonstrated measurable effects, but can shrink in real-world feeds, fade without reinforcement, and be overwhelmed by competing content. Cultural adaptation is essential and underinvested, and the 2028 cycle in particular will demand capabilities most organizations have not yet built.

The smart and careful use of inoculation must be included in campaigns and organizational communications strategies as a central component. We now live in an information environment that is structured in ways that allow disinformation to spread quickly and effectively.

The organizations that start building community relationships, AI-assisted pipelines, cross-sector coordination, sustained campaigns, and built-in measurement *now* will be best positioned to protect democratic resilience in the years ahead.

Glossary of Terms

Accuracy prompt: A brief intervention that asks people to consider whether information is accurate before deciding whether to share it. Accuracy prompts are intended to shift attention toward truthfulness at the moment of decision.

Booster: A follow-up exposure to an inoculation intervention designed to refresh or reinforce its effects after they begin to fade.

Conspiracy theory: An explanation that attributes events to secret plots by powerful or malicious actors, often without adequate supporting evidence. In inoculation research, conspiracy framing is studied as a recurring manipulation technique.

Discernment: The ability to distinguish reliable or accurate information from false, misleading, or manipulative information.

Ecological validity: The extent to which the results of a study are likely to apply under real-world conditions rather than only in a controlled experimental setting.

Fake experts: People presented as credible authorities despite lacking relevant expertise, or purported experts used to create a misleading impression of scientific or professional support.

Field experiment / field study: Research conducted in a real-world environment rather than solely in a controlled experimental setting. In this report, this can include interventions delivered to people through actual social-media platforms or community settings.

Hierarchical Bayesian modeling: A statistical approach that looks for overall patterns across many studies or groups while also accounting for differences among individual studies and participants.

Inoculation: A strategy for building resistance to misinformation by exposing people in advance to weakened examples of misleading arguments or manipulation techniques and explaining how they work.

In-the-wild study: An informal term for research testing an intervention under actual real-world conditions, such as within a live social-media platform, rather than in a controlled or simulated environment.

Logic-based prebunking: Prebunking that teaches people to recognize flaws in reasoning, such as whataboutism, moving the goalposts, or strawman arguments.

Media literacy: The knowledge and skills needed to critically evaluate, interpret, and engage with information and media.

Moving the goalposts: Changing the standard of proof or requirements of an argument after the original standard has been met.

Polarization: In the context of manipulation, using divisive rhetoric to exaggerate differences between groups and encourage an “us versus them” view.

Prebunking: An intervention delivered **before** people encounter misinformation that warns them about misleading claims or manipulation techniques and helps them recognize and resist them.

Psychological inoculation: The theoretical principle underlying prebunking: just as exposure to a weakened pathogen can build biological resistance, advance exposure to weakened misleading arguments or manipulation techniques can build resistance to later persuasion.

Refutational prebunking: A form of prebunking that anticipates a particular false or misleading claim and provides information refuting it before people encounter the misinformation itself.

Signal detection theory: A framework for distinguishing genuine improvement in people's ability to tell true from false information from a general tendency to become more skeptical of *all* information.

Simulated social-media feed: An experimental environment constructed to resemble a real social-media feed, allowing researchers to test interventions under more realistic information conditions while retaining experimental control.

Strawman fallacy: Misrepresenting or oversimplifying another person's argument so the distorted version is easier to attack.

Technique-based prebunking: Prebunking that teaches people to recognize recurring manipulation tactics, such as emotional manipulation, polarization, scapegoating, or fake experts, so they can identify those techniques when they appear in new claims or contexts.

Trusted messenger: A person or organization whose existing credibility or relationship with an audience may make information more persuasive or easier to apply. Trusted messengers can include community organizations and other locally credible sources.

Whataboutism: Deflecting criticism or an argument by pointing to someone else's alleged wrongdoing rather than addressing the original issue.

References

- Priyadarshi Amar, Sumitra Badrinathan, Simon Chauchard and Florian Sichart, “Countering misinformation early: Evidence from a classroom-based field experiment in India,” American Political Science Review, 2026. https://doi.org/10.1017/S0003055425101184
- Frederico Batista Pereira, Natália S. Bueno, Felipe Nunes and Nara Pavão, "Inoculation reduces misinformation: Experimental evidence from multidimensional interventions in Brazil," Journal of Experimental Political Science, 2024. https://doi.org/10.1017/XPS.2023.11
- Mikey Biddlestone and colleagues, "Video inoculation against election misinformation across 12 EU nations," Communications Psychology, 2026. https://doi.org/10.1038/s44271-025-00379-3
- Mikey Biddlestone, Jon Roozenbeek, Jane Suiter, Eileen Culloty and Sander van der Linden, "Tune in to the prebunking network! Development and validation of six inoculation videos that prebunk manipulation tactics and logical fallacies in misinformation," Political Psychology, 2025. https://doi.org/10.1111/pops.70015
- Esther Boissin, Thomas H. Costello, Daniel Spinoza-Martín, David G. Rand and Gordon Pennycook, "Dialogues with large language models reduce conspiracy beliefs even when the AI is perceived as human," PNAS Nexus, 2025. https://doi.org/10.1093/pnasnexus/pgaf325
- Georgia Capewell, Rakoen Maertens, Miriam Remshard, Sander van der Linden, Josh Compton, Stephan Lewandowsky and Jon Roozenbeek, "Misinformation interventions decay rapidly without an immediate posttest," Journal of Applied Social Psychology, 2024. https://doi.org/10.1111/jasp.13049
- John M. Carey, Brian Fogarty, Marília Gehrke, Brendan Nyhan and Jason Reifler, "Prebunking and credible source corrections increase election credibility: Evidence from the US and Brazil," Science Advances, 2025. https://doi.org/10.1126/sciadv.adv3758
- Thomas H. Costello, Gordon Pennycook and David G. Rand, "Durably reducing conspiracy beliefs through dialogues with artificial intelligence," Science, 2024. https://doi.org/10.1126/science.adq1814
- European Fact-Checking Standards Network (EFCSN), “Adding to the Fact-Checking Toolkit: Prebunking,” 2025.
- Kobi Hackenburg, Ben M. Tappin, Luke Hewitt and colleagues, "The levers of political persuasion with conversational artificial intelligence," Science, 2025. https://doi.org/10.1126/science.aea3884
- Emma Hoes, Brian Aitken, Jingwen Zhang, Tomasz Gackowski and Magdalena Wojcieszak, "Prominent misinformation interventions reduce misperceptions but increase scepticism," Nature Human Behaviour, 2024. https://doi.org/10.1038/s41562-024-01884-x

- Guanxiong Huang, Wufan Jia and Wenting Yu, "Media literacy interventions improve resilience to misinformation: A meta-analytic investigation of overall effect and moderating factors," Communication Research, 2024. https://doi.org/10.1177/00936502241288103
- Angela Y. Lee, Ryan C. Moore and Jeffrey T. Hancock, “Building resilience to misinformation in communities of color: Results from two digital media literacy interventions,” New Media & Society, 2025. https://doi.org/10.1177/14614448241227841
- Hause Lin, Gabriela Czarnek, Benjamin Lewis, Joshua P. White, Adam J. Berinsky, Thomas Costello, Gordon Pennycook and David G. Rand, “Persuading voters using human-artificial intelligence dialogues,” Nature, 2025. https://doi.org/10.1038/s41586-025-09771-9
- Mitchell Linegar, Betsy Sinclair, Sander van der Linden and R. Michael Alvarez, "Towards scalable AI-assisted pre-bunking of election misinformation," Royal Society Open Science, 2026. https://doi.org/10.1098/rsos.252226
- John A. List, Lina M. Ramírez, Julia Seither, Jaime Unda and Beatriz H. Vallejo, "Critical thinking and misinformation vulnerability: Experimental evidence from Colombia," PNAS Nexus, 2024. https://doi.org/10.1093/pnasnexus/pgae361
- Daniel Loughnan, Aart van Stekelenburg, J. Loes Pouwels and Mariska Kleemans, "Concerns regarding the methodology of a psychological inoculation meta-analysis on misinformation," Journal of Medical Internet Research, 2025. https://doi.org/10.2196/64430
- Chang Lu, Bo Hu, Qiang Li, Chao Bi and Xing-Da Ju, "Psychological inoculation for credibility assessment, sharing intention, and discernment of misinformation: Systematic review and meta-analysis," Journal of Medical Internet Research, 2023. https://doi.org/10.2196/49255
- Rakoen Maertens, Jon Roozenbeek, Jon S. Simons, Stephan Lewandowsky, Vanessa Maturo, Beth Goldberg, Rachel Xu and Sander van der Linden, “Psychological booster shots targeting memory increase long-term resistance against misinformation,” Nature Communications, 2025. https://doi.org/10.1038/s41467-025-57205-x
- Cameron Martel and colleagues, "On the efficacy of accuracy prompts across partisan lines: An adversarial collaboration," Psychological Science, 2024. https://doi.org/10.1177/09567976241232905
- Ariana Modirrousta-Galian and Philip A. Higham, "Gamified inoculation interventions do not improve discrimination between true and fake news: Reanalyzing existing research with receiver operating characteristic analysis," Journal of Experimental Psychology: General, 2023. https://doi.org/10.1037/xge0001395
- Thomas Nygren and Evgenia Efimova, "Investigating the long-term impact of misinformation interventions in upper secondary education," PLOS ONE, 2025. https://doi.org/10.1371/journal.pone.0326928
- Thomas Nygren, Emily R. Spearing, Nicolas Fay, Davide Vega, Isabella I. Hardwick, Jon Roozenbeek and Ullrich K. H. Ecker, "The seven roles of generative AI: Potential

and pitfalls in combatting misinformation," Behavioral Science & Policy, 2026. https://doi.org/10.1177/23794607261417815
- Molly Offer-Westort, Leah R. Rosenzweig and Susan Athey, "Battling the coronavirus 'infodemic' among social media users in Kenya and Nigeria," Nature Human Behaviour, 2024. https://doi.org/10.1038/s41562-023-01810-7
- Gordon Pennycook, David G. Rand, Jon Roozenbeek, Sander van der Linden, Ziv Epstein, Mohsen Mosleh, Antonio A. Arechar and Dean Eckles, “Inoculation and accuracy prompting increase accuracy discernment in combination but not alone,” Nature Human Behaviour, 2024. https://doi.org/10.1038/s41562-024-02023-2
- Jon Roozenbeek, Jana Lasser, Malia Marks, Tianzhu Qin, David Garcia, Beth Goldberg, Ramit Debnath, Sander van der Linden and Stephan Lewandowsky, “Misinformation interventions and online sharing behaviour: Lessons learned from two pre-registered field studies,” Royal Society Open Science, 2025. https://doi.org/10.1098/rsos.251377
- Jon Roozenbeek, Miriam Remshard and Yara Kyrychenko, "Beyond the headlines: On the efficacy and effectiveness of misinformation interventions," Advances in Psychology, 2024. https://doi.org/10.56296/aip00019
- Nabil Saleh, Fadi Makki, Sander van der Linden and Jon Roozenbeek, "Inoculating against extremist persuasion techniques: Results from a randomized controlled trial in post-conflict areas in Iraq," Advances in Psychology, 2023. https://doi.org/10.56296/aip00005
- Tina Seabrooke, Ariana Modirrousta-Galian and Philip A. Higham, "Re-examining the bad news game: No evidence of improved discrimination of Indian true and fake news headlines," Psychonomic Bulletin & Review, 2025. https://doi.org/10.3758/s13423-025-02827-x
- Charles Senteio and colleagues, "Overcoming health misinformation in marginalized groups: A systematic review," International Journal for Equity in Health, 2025. https://doi.org/10.1186/s12939-025-02657-2
- Almog Simchon, Tomer Zipori, Louis Teitelbaum, Stephan Lewandowsky and Sander van der Linden, "A signal detection theory meta-analysis of psychological inoculation against misinformation," Current Opinion in Psychology, 2026. https://doi.org/10.1016/j.copsyc.2025.102194
- H. Holden Thorp, “Editorial Expression of Concern,” Science, 2026. https://doi.org/10.1126/science.aej2383
- Cecilie S. Traberg, Jon Roozenbeek and Sander van der Linden, "Gamified inoculation reduces susceptibility to misinformation from political in-groups," Harvard Kennedy School Misinformation Review, 2024. https://misinforeview.hks.harvard.edu/article/gamified-inoculation-reduces-susceptibility-to-misinformation-from-political-ingroups/
- Sander van der Linden, Debra Louison-Lavoy, Nicholas Blazer, Nancy S. Noble and Jon Roozenbeek, "Prebunking misinformation techniques in social media feeds: Results

from an Instagram field study," Harvard Kennedy School Misinformation Review, 2026. https://doi.org/10.37016/mr-2020-193
- Sze Yuh Nina Wang, Samantha C. Phillips, Kathleen M. Carley, Hause Lin and Gordon Pennycook, "Limited effectiveness of psychological inoculation against misinformation in a social media feed," PNAS Nexus, 2025. https://doi.org/10.1093/pnasnexus/pgaf172
- *Preprint (not peer-reviewed):* Authors pending confirmation, "The impact of prebunking interventions against misinformation on discrimination ability and criterion: An individual-participant-data network meta-analysis," 2025, Research Square. https://doi.org/10.21203/rs.3.rs-6660774/v1

Endnotes

[1] Mikey Biddlestone et al., “Video inoculation against election misinformation across 12 EU nations,” Communications Psychology, 2026; European Fact-Checking Standards Network, “Adding to the fact-checking toolkit: Prebunking,” 2025.
[2] Guanxiong Huang, Wufan Jia and Wenting Yu, “Media literacy interventions improve resilience to misinformation: A meta-analytic investigation of overall effect and moderating factors,” Communication Research, 2024. https://doi.org/10.1177/00936502241288103
[3] John M. Carey, Brian Fogarty, Marília Gehrke, Brendan Nyhan and Jason Reifler, “Prebunking and credible source corrections increase election credibility: Evidence from the US and Brazil,” Science Advances, 2025. https://doi.org/10.1126/sciadv.adv3758
[4] Cameron Martel and colleagues, "On the efficacy of accuracy prompts across partisan lines: An adversarial collaboration," Psychological Science, 2024. https://doi.org/10.1177/09567976241232905
[5] Rakoen Maertens, Jon Roozenbeek, Jon S. Simons, Stephan Lewandowsky, Vanessa Maturo, Beth Goldberg, Rachel Xu and Sander van der Linden, “Psychological booster shots targeting memory increase long-term resistance against misinformation,” Nature Communications, vol. 16, Article 2062 (2025). https://doi.org/10.1038/s41467-025-57205-x
[6] John M. Carey, Brian Fogarty, Marília Gehrke, Brendan Nyhan and Jason Reifler, "Prebunking and credible source corrections increase election credibility: Evidence from the US and Brazil," Science Advances, 2025. https://doi.org/10.1126/sciadv.adv3758
[7] Cecilie S. Traberg, Jon Roozenbeek and Sander van der Linden, "Gamified inoculation reduces susceptibility to misinformation from political in-groups," Harvard Kennedy School Misinformation Review, 2024. https://misinforeview.hks.harvard.edu/article/gamified-inoculation-reduces-susceptibility-to-misinformation-from-political-ingroups/
[8] Biddlestone et al., “Video inoculation against election misinformation across 12 EU nations.”
[9] Sander van der Linden, Debra Louison-Lavoy, Nicholas Blazer, Nancy S. Noble and Jon Roozenbeek, "Prebunking misinformation techniques in social media feeds: Results from an Instagram field study," Harvard Kennedy School Misinformation Review, 2026. https://doi.org/10.37016/mr-2020-193
[10] Mitchell Linegar, Betsy Sinclair, Sander van der Linden and R. Michael Alvarez, “Towards scalable AI-assisted pre-bunking of election misinformation,” *Royal Society Open Science*, 2026. https://doi.org/10.1098/rsos.252226
[11] Carey et al., "Prebunking and credible source corrections increase election credibility."
[12] Guanxiong Huang, Wufan Jia and Wenting Yu, "Media literacy interventions improve resilience to misinformation: A meta-analytic investigation of overall effect and moderating factors," Communication Research, 2024. https://doi.org/10.1177/00936502241288103

[13] Maertens et al., “Psychological booster shots targeting memory.”
[14] van der Linden et al., “Prebunking misinformation techniques in social media feeds.”
[15] van der Linden et al., "Prebunking misinformation techniques in social media feeds."
[16] Mitchell Linegar, Betsy Sinclair, Sander van der Linden and R. Michael Alvarez, “Towards scalable AI-assisted pre-bunking of election misinformation,” Royal Society Open Science, 2026. https://doi.org/10.1098/rsos.252226
[17] Kobi Hackenburg, Ben M. Tappin, Luke Hewitt and colleagues, “The levers of political persuasion with conversational artificial intelligence,” Science, 2025. https://doi.org/10.1126/science.aea3884
[18] Priyadarshi Amar, Sumitra Badrinathan, Simon Chauchard and Florian Sichart, "Countering misinformation early: Evidence from a classroom-based field experiment in India," American Political Science Review 120, no. 2 (2026): 725–745. https://doi.org/10.1017/S0003055425101184
[19] Angela Y. Lee, Ryan C. Moore and Jeffrey T. Hancock, “Building resilience to misinformation in communities of color: Results from two digital media literacy interventions,” New Media & Society, vol. 27, no. 6 (2025), 3545–3576.
[20] Carey et al., "Prebunking and credible source corrections increase election credibility."
[21] Biddlestone et al., “Video inoculation against election misinformation across 12 EU nations.”
[22] van der Linden et al., "Prebunking misinformation techniques in social media feeds."
[23] Linegar et al., "Towards scalable AI-assisted pre-bunking of election misinformation."
[24] Hause Lin, Gabriela Czarnek, Benjamin Lewis, Joshua P. White, Adam J. Berinsky, Thomas Costello, Gordon Pennycook and David G. Rand, "Persuading voters using human-artificial intelligence dialogues," Nature 648 (2025): 394–401. https://doi.org/10.1038/s41586-025-09771-9
[25] Thomas H. Costello, Gordon Pennycook and David G. Rand, "Durably reducing conspiracy beliefs through dialogues with artificial intelligence," Science, 2024. https://doi.org/10.1126/science.adq1814
[26] Esther Boissin, Thomas H. Costello, Daniel Spinoza-Martín, David G. Rand and Gordon Pennycook, "Dialogues with large language models reduce conspiracy beliefs even when the AI is perceived as human," PNAS Nexus, 2025. https://doi.org/10.1093/pnasnexus/pgaf325
[27] H. Holden Thorp, “Editorial Expression of Concern,” Science, vol. 392, no. 6803 (2026), 1131. https://doi.org/10.1126/science.aej2383
[28] Martel et al., "On the efficacy of accuracy prompts across partisan lines."
[29] Sze Yuh Nina Wang, Samantha C. Phillips, Kathleen M. Carley, Hause Lin and Gordon Pennycook, "Limited effectiveness of psychological inoculation against misinformation in a social media feed," PNAS Nexus, 2025. https://doi.org/10.1093/pnasnexus/pgaf172
[30] Jon Roozenbeek, Jana Lasser, Malia Marks, Tianzhu Qin, David Garcia, Beth Goldberg, Ramit Debnath, Sander van der Linden and Stephan Lewandowsky, “Misinformation interventions and online sharing behaviour: Lessons learned from two pre-registered field studies,” Royal Society Open Science, 2025. https://doi.org/10.1098/rsos.251377
[31] Jon Roozenbeek, Miriam Remshard and Yara Kyrychenko, "Beyond the headlines: On the efficacy and effectiveness of misinformation interventions," Advances in Psychology, 2024. https://doi.org/10.56296/aip00019
[32] Thomas Nygren and Evgenia Efimova, "Investigating the long-term impact of misinformation interventions in upper secondary education," PLOS ONE, 2025. https://doi.org/10.1371/journal.pone.0326928
[33] Georgia Capewell, Rakoen Maertens, Miriam Remshard, Sander van der Linden, Josh Compton, Stephan Lewandowsky and Jon Roozenbeek, "Misinformation interventions decay rapidly without an immediate posttest," Journal of Applied Social Psychology 54, no. 8 (2024): 441–454. https://doi.org/10.1111/jasp.13049
[34] Maertens et al., “Psychological booster shots targeting memory.”
[35] Lee, Moore and Hancock, "Building resilience to misinformation in communities of color."
[36] Onyx Impact, "Aisha: An Immersive App for Understanding AI Bias," The Atlanta Voice, June 25, 2026. https://theatlantavoice.com/onyx-impact-launches-aisha/
[37] Mikey Biddlestone, Jon Roozenbeek, Jane Suiter, Eileen Culloty and Sander van der Linden, “Tune in to the prebunking network! Development and validation of six inoculation videos that prebunk

manipulation tactics and logical fallacies in misinformation," Political Psychology, 2025. https://doi.org/10.1111/pops.70015
[38] Almog Simchon, Tomer Zipori, Louis Teitelbaum, Stephan Lewandowsky and Sander van der Linden, "A signal detection theory meta-analysis of psychological inoculation against misinformation," Current Opinion in Psychology, 2026. https://doi.org/10.1016/j.copsyc.2025.102194
[39] Huang, Jia and Yu, "Media literacy interventions improve resilience to misinformation."
[40] Huang, Jia and Yu, "Media literacy interventions improve resilience to misinformation."
[41] Biddlestone et al., "Tune in to the prebunking network."
[42] Martel et al., "On the efficacy of accuracy prompts across partisan lines."
[43] Gordon Pennycook et al., "Inoculation and accuracy prompting increase accuracy discernment in combination but not alone," Nature Human Behaviour, 2024. https://doi.org/10.1038/s41562-024-02023-2
[44] Capewell et al., "Misinformation interventions decay rapidly without an immediate posttest."
[45] Maertens et al., "Psychological booster shots targeting memory."
[46] Huang, Jia and Yu, "Media literacy interventions improve resilience to misinformation."
[47] Linegar et al., "Towards scalable AI-assisted pre-bunking of election misinformation."
[48] Thomas Nygren, Emily R. Spearing, Nicolas Fay, Davide Vega, Isabella I. Hardwick, Jon Roozenbeek and Ullrich K. H. Ecker, "The seven roles of generative AI: Potential and pitfalls in combatting misinformation," Behavioral Science & Policy, 2026. https://doi.org/10.1177/23794607261417815
[49] Linegar et al., "Towards scalable AI-assisted pre-bunking of election misinformation."