



Countering Antisemitic Manipulation Online

Two randomized studies with more than 6,000 U.S. adults were conducted, testing whether short “prebunking” (or “inoculation”) videos teach people to resist antisemitic manipulation. The study tested how long the inoculation effect lasts, whether the ability to recognize manipulation techniques carries to new content, and whether an AI presenter can deliver an inoculation message as well as a human creator.

6,000+

U.S. adults across two randomized studies

72%

Caught the manipulation in a brand-new post

1 month+

Protection still measurable at 4½ weeks

4

Antisemitic narratives the approach was tested against

78,700

Views reached through trusted online creators

WHY THIS MATTERS

Antisemitic narratives spread across social media faster than the tools (e.g., content moderation, fact-checking, and debunking) built to counter them can respond. By the time a false claim is corrected, millions have already seen it. **Psychological inoculation, or “prebunking,” takes a different approach:** instead of reacting after a harmful antisemitic narrative has spread, it teaches people to recognize the manipulation technique *before* they encounter it, much like a vaccine prepares the body to fight an infectious disease before exposure. The goal is not to tell people what to think, but to strengthen the critical-thinking skills that enable them to spot manipulation on their own.

This study set out to answer the practical questions that determine whether this approach can protect people at scale: does it actually work, how long does the protection last, does the ability to recognize manipulation techniques transfer to new tropes, and, crucially for cost and speed, can an AI avatar deliver an inoculation message as effectively as a human creator can?

THE RESEARCH QUESTION

Can short prebunking videos teach people to resist the manipulation techniques used in antisemitic narratives and can an AI avatar deliver the inoculation message as well as a real person?

WHAT IS PREBUNKING?

Prebunking gives viewers a small, clearly labeled dose of a manipulation technique together with a plain explanation of how it works so they can recognize and resist the technique when they meet it in the wild.

Because it targets **tactics rather than beliefs**, the approach is nonpartisan and works across communities and narratives. It teaches a reusable skill: once you learn how to spot a manipulation technique, you can catch it in content you have not yet encountered. Since the same manipulation techniques are used across different types of narratives, prebunking against antisemitism has the added benefit of inoculating people against related harms (e.g., white supremacy, violent misogyny, etc.).

HOW WE TESTED IT

PERIL ran two randomized studies on the Prolific research platform. People were randomly assigned to watch short inoculation videos (the test groups) or nothing at all (the comparison groups), then the groups were tested to see if they could recognize the manipulation techniques being used in four antisemitic tropes and whether they could apply that recognition skill to new posts unrelated to antisemitism.

Study 1 followed 2,394 adults through **three periods**: the initial test, a follow-up test about two weeks later, and a final test about 4½ weeks later. The tests focused on the detection of Holocaust distortion techniques, comparing human vs. AI presenters, and 20- vs. 45-second versions of the video. **Study 2** extended the approach to three more narratives with 3,698 adults.

THE FOUR NARRATIVES AND THE MANIPULATION TACTIC EACH VIDEO TEACHES

Holocaust Denial & Distortion

Claims that the Holocaust was exaggerated or fabricated.

TACTIC TAUGHT

Manufactured doubt: raising misleading questions to undermine settled facts.

Jewish Control Conspiracies

Claims that Jews secretly control media, government, or the economy.

TACTIC TAUGHT

Scapegoating: blaming a single group for complex societal problems.

Great Replacement

Claims that Jews orchestrate demographic change through immigration.

TACTIC TAUGHT

False causation: pinning broad social trends on one group.

Jews as Inherently Violent

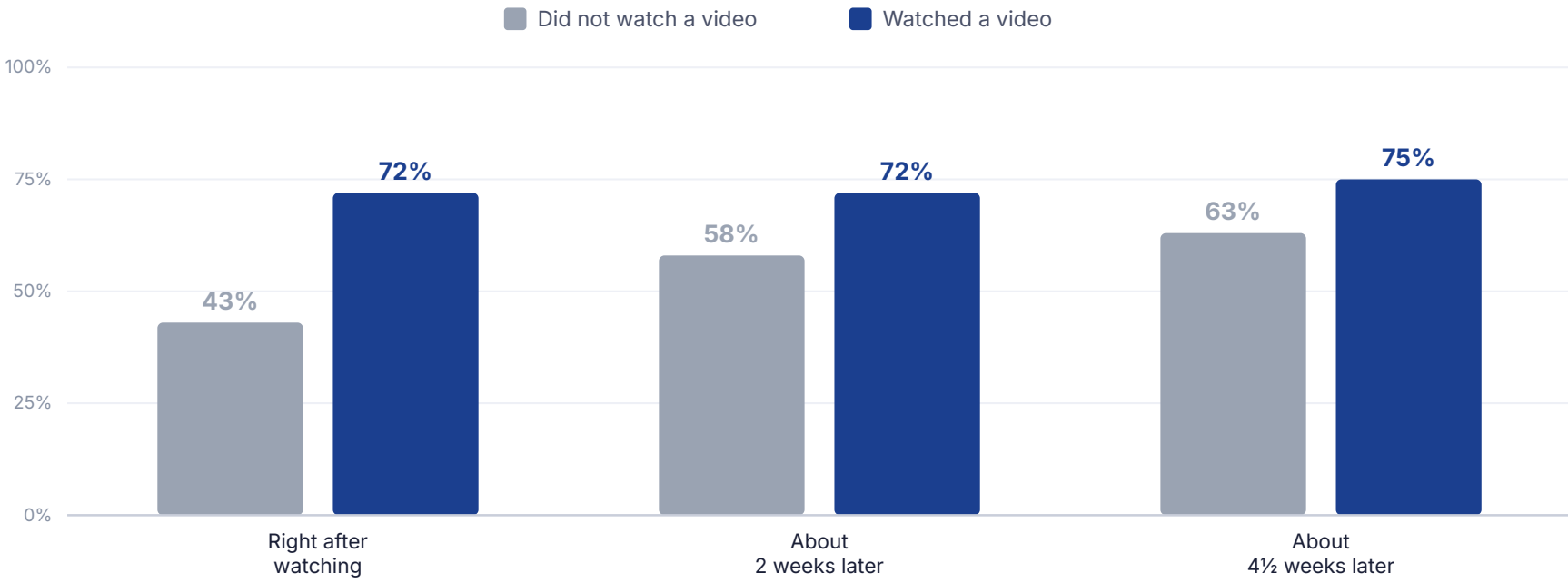
Claims holding all Jews responsible for the actions of the Israeli government.

TACTIC TAUGHT

Collective blame: holding a whole group responsible for the acts of some.

THE VIDEOS TAUGHT A REAL, REUSABLE SKILL AND IT LASTED

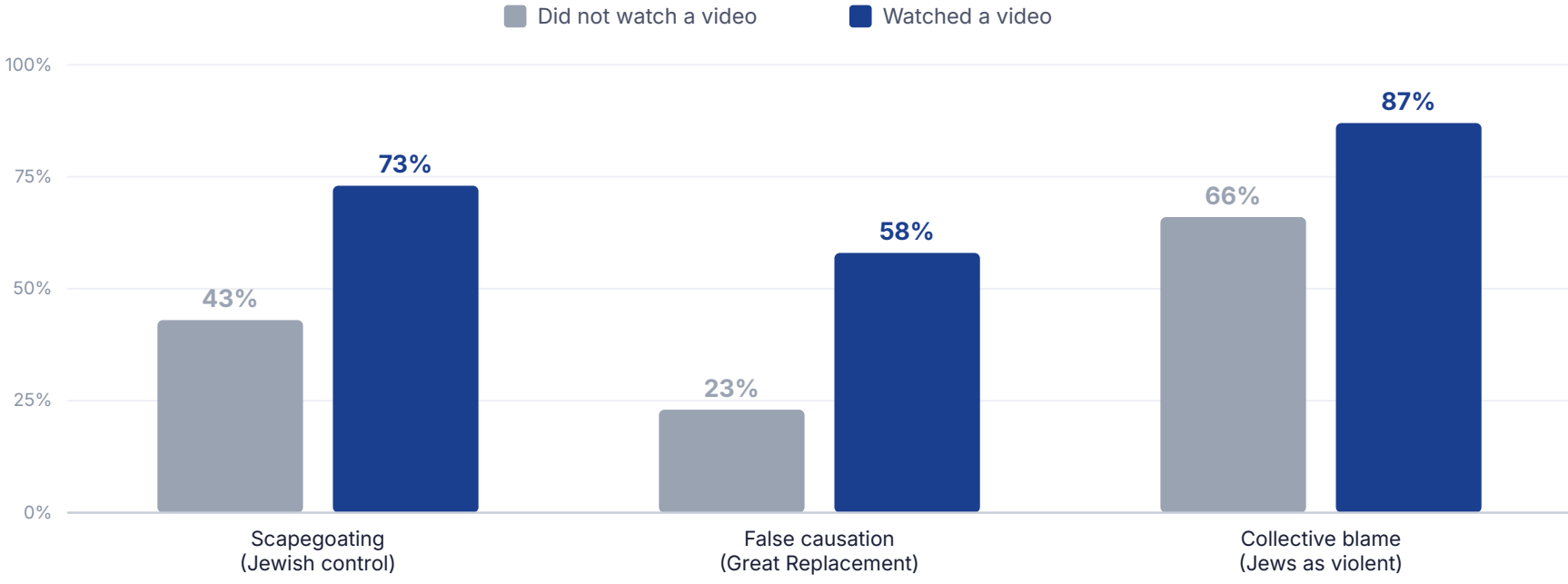
After watching a single short video, people were far better at spotting the same manipulation technique in a different post that was completely unrelated to antisemitism: **72% caught it versus 43%** of people who had not watched the video. That gap means viewers learned a transferable skill, not just how to recognize manipulation in one antisemitic narrative. This suggests that exposing manipulation in one type of hateful narrative has the potential to expose its use in any type of hateful narrative, not just antisemitism. We also found that the inoculation effect held up. We checked back at two weeks and 4½ weeks after the initial test, and video-watchers remained better at recognizing the manipulation technique than the control group.



How to read this: Each pair shows the share of people who correctly recognized the manipulation technique in a fresh, unrelated post. People who watched the prebunking video (blue) remained better at spotting the technique well after viewing the video. Non-watchers (gray) improved too but video watchers stayed clearly ahead throughout. Even people who could no longer remember watching the video still outperformed those who never saw one.

IT WORKED ACROSS EVERY NARRATIVE TESTED

The same approach taught people to catch the manipulation behind several distinct narratives. Some tactics were easier to teach than others: scapegoating was the easiest, false causation was the hardest. But watching a prebunking video produced a clear improvement in manipulation recognition for **every one**.



How to read this: For each narrative, the share of people who caught the manipulation technique in a fresh, unrelated post. In every case, people who watched a prebunking video (blue) were much better at catching it than those who did not watch a video (gray). A stricter check confirmed this was a genuine skill: viewers spotted real manipulation without simply calling everything manipulative.

An AI Presenter Taught Manipulation Recognition as Well as a Person

We compared videos delivered by a human creator with ones delivered by an AI avatar (that was clearly labeled “AI Generated.”). The **AI version taught manipulation recognition just as well** as the version featuring a human creator. Viewers liked and trusted the AI presenter noticeably less, yet learned just as much from the video, which tells us the benefit comes from the content, not the messenger. Because AI videos are far cheaper and faster to produce and refresh, this presents a practical path to create content at scale.

The Ability to Recognize the Manipulation Technique Outlives the Memory of Watching the Video

About 4½ weeks later, 57% of viewers could no longer recall the video they had watched, yet that same group was still better at recognizing the manipulation technique than the group that didn’t watch the prebunking video. Inoculation is not tied to remembering the video; the underlying skill persists even after the video itself is forgotten. The inoculation effect fades from its early peak but levels off rather than disappearing, which suggests that the occasional use of “booster” videos can make the inoculation effect last over long periods.

ONE IMPORTANT CAUTION

Testing videos before they are released allows us to catch any backfire effects, and here it did. The video addressing the **“collective blame” narrative** made the trope seem slightly more credible to people who watched the prebunking video. This finding was confirmed after a retest of the video and a test of the video using a slightly altered script. Separately, the shortest AI video (about 20 seconds) taught manipulation recognition well but it did not hold up as well over time. Short, AI-generated videos need more testing before scaling. **No other video showed any harmful effect at any point.**

REACHING REAL AUDIENCES THROUGH TRUSTED CREATORS

The four prebunking videos that passed testing were shared by trusted creators on TikTok and Instagram and audiences did not just watch them, they shared them in their networks.

78,700

Views across TikTok & Instagram

749 hrs

Total time spent watching

1,190

Shares & reposts beyond the original audience

WHAT THIS MEANS FOR INOCULATION MOVING FORWARD

Prebunking Works and Is Ready to Scale

It builds a **durable, reusable skill**, not a one-time reaction. This skill transfers across multiple antisemitic narratives, and demographic and ideological groups.

AI-Made Videos Are a Cost-Effective Path

At the standard **45-second length**, AI-delivered videos teach manipulation recognition as well as human creators, at a fraction of the cost and time to produce and refresh.

Transferability

Prebunking teaches manipulation recognition across contexts, meaning that inoculating against one type of hate helps inoculate against all types of hate.



boundless

This factsheet presents the findings of a partnership between PERIL and Boundless exploring the effectiveness of video-based inoculation in combating growing antisemitism online. It was conducted from January to July 2026.