



Association of
Internet Research
Specialists

AOFIRS RESEARCH REPORT

DIGITAL VERIFICATION
MEDIA LITERACY
ONLINE TRUST

The Visual Fact Check

How to Spot Deepfakes, Sponsored Posts, and
Digital Fakes Without Specialized Forensic
Software Tools



DEEPPAKES



VISUAL CLUES



- ✓ Unnatural lighting
- ✓ Odd facial details
- ✓ Inconsistent hands/text



SPONSORED POSTS



- \$ Sponsored labels
- 🚫 Missing disclosure
- 📊 Unusual engagement patterns



SOURCE CHECKS



- 🌐 Reverse image search
- 🔗 Check the source
- 📄 Compare versions



RED FLAGS

- ⚠️ Sensational claims
- 👤 Unverified accounts
- 🔄 Recycled media



VISUAL CLUES

- Unnatural lighting
- Odd facial details
- Inconsistent hands/text



PLATFORM SIGNALS

- Sponsored labels
- Missing disclosure
- Engagement patterns



VERIFICATION STEPS

- Reverse image search
- Cross-check the source
- Compare versions



WHY IT MATTERS

- Trust
- Informed decisions
- Safer online research

The Visual Fact Checks

Detecting Fake News, Deepfakes and Hidden Persuasion

Logical Verification Methods for Internet Researchers: Evidence,
Experiments, Bias and Practical Countermeasures

A comprehensive research report on detecting fake news, deciphering deepfakes through logical verification, recognizing advertising hidden in content, and identifying and countering bias, drawing on AOFIRS Lesson 31 (Search in News, Articles & Blogs).

Association of Internet Research Specialists (AOFIRS)

October 2026

By: Naveed Manzoor

© 2026 Association of Internet Research Specialists (AOFIRS). All rights reserved.

Copyright Statement

Copyright © 2026 Association of Internet Research Specialists (AOFIRS). All rights reserved.

All rights in this report, including its text, original tables, charts, diagrams, frameworks, checklists and index, belong exclusively to the Association of Internet Research Specialists (AOFIRS). No part of this publication may be reproduced, distributed, stored in a retrieval system, adapted, translated, or transmitted in any form or by any means, whether electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of AOFIRS.

Brief quotations may be used in reviews, teaching and scholarly works, provided that AOFIRS is clearly acknowledged as the source and the use is consistent with applicable copyright law.

Third-party works, data, trademarks and names cited in this report remain the property of their respective owners and are referenced for research, commentary and educational purposes. Charts and diagrams were created by AOFIRS using data published in the cited sources; their inclusion does not imply endorsement by those sources.

Disclaimer. This report is provided for educational and research purposes only. While every effort has been made to verify facts and figures against their original sources as of October 2026, online platforms, search operators and regulations change frequently. AOFIRS accepts no liability for decisions made on the basis of this material. Nothing in this report constitutes legal advice.

Suggested citation (APA 7)

Association of Internet Research Specialists. (2026). *Detecting fake news, deepfakes and hidden persuasion: Logical verification methods for internet researchers* [Research report]. AOFIRS.

Companion publication

This report is accompanied by an ancillary general-audience article, "Seeing Isn't Believing: A Plain-Language Guide to Spotting Fake News, Deepfakes and Hidden Ads," also published by AOFIRS.

Table of Contents

Executive Summary	1
Key Findings	1
1. Introduction	3
1.1 Why This Report Matters.....	3
1.2 Scope and Research Questions.....	3
1.3 How the Evidence Was Gathered	3
2. Key Concepts and Definitions	5
2.1 Misinformation, Disinformation and Malinformation	5
2.2 Deepfakes and Cheap Fakes	5
2.3 Hidden Advertising	5
2.4 Astroturfing, Front Groups and Manufactured Doubt	6
3. Why False Content Travels So Well.....	7
3.1 Novelty and Emotion	7
3.2 Inattention Rather Than Ideology.....	7
3.3 Repetition Breeds Belief	7
3.4 Who Shares, and Where People Now Get News	7
3.5 The Machine-Made Content Surge.....	8
4. Detecting Fake News: Methods That Work.....	10
4.1 Read Laterally, Not Vertically	10
4.2 The SIFT Moves.....	10
4.3 Why Checklists Alone Fall Short.....	11
4.4 Using Search Operators to Find the Original Story	11
4.5 Reading the Clock: Timestamps, Datelines and Archives	12
4.6 Escaping the Local Bubble: Region and Language	13
4.7 Transparency Signals and Newswires	13
5. Deciphering Deepfakes Without Complicated Software	14
5.1 What the Experiments Tell Us	14
5.2 A Layered Approach.....	15
5.3 Layer 1: Surface Artifacts	16
5.4 Layer 2: Context and Logic.....	17
5.5 Layer 3: Provenance and Corroboration.....	17
5.6 Layer 4: Out-of-Band Confirmation	18
5.7 The Liar's Dividend.....	18
6. Spotting Advertising Hidden in Content.....	19
6.1 The Scale of the Problem	19
6.2 Common Formats	20
6.3 A Quick Test for Hidden Persuasion	20
6.4 What the Rules Require	21
7. How Murmurs, Marketers and Lobbyists Distort Impressions.....	22
7.1 Social Media Murmurs and Coordinated Campaigns	22
7.2 Front Groups and Think-Tank Branding.....	22
7.3 Manufactured Doubt	22

7.4 News Influencers and Blurred Lines	23
8. Experiments, Test Cases and Statistics.....	24
8.1 Summary of Key Experiments.....	24
8.2 What Interventions Achieve	25
8.3 Real-World Test Cases	26
8.4 Search Test Cases from Lesson 31	26
8.5 Reflecting on the Statistics	27
9. The Problem of Verification.....	28
9.1 What "Verified" Should Mean	28
9.2 Verification Methods Compared	28
9.3 Practical Obstacles.....	29
9.4 Correcting Without Backfiring	29
10. Types of Bias and How to Counter Them	30
10.1 Biases in the Researcher	30
10.2 Biases in Sources and Media.....	30
10.3 Biases in Search Tools.....	30
10.4 A Researcher's Debiasing Routine	31
11. The Researcher's Practical Toolkit	33
12. Limitations of This Report	35
13. Conclusion	36
Appendix A: Query Templates for Verification	37
References.....	38
Index	44

List of Figures

Figure 1. Share of people concerned about what is real and fake online, 2022–2026	8
Figure 2. Growth of AI-generated "content farm" news sites tracked by NewsGuard	9
Figure 3. Lateral reading before and after a short online course (n = 87)	10
Figure 4. Human accuracy at identifying synthetic media across major studies	15
Figure 5. Four layers of logical deepfake verification	16
Figure 6. Recognition and disclosure of hidden advertising across studies	19
Figure 7. Perceived reliability of posts before and after playing the "Bad News" game	25
Figure 8. The seven-step logical verification loop	33

List of Tables

Table 1. Three types of information disorder	5
Table 2. Search-operator patterns for tracing a news story to its source	11
Table 3. Kinds of date a researcher may encounter	12
Table 4. Surface checks for images, video and audio	16
Table 5. Formats of hidden advertising and how to recognize them	20
Table 6. Who distorts impressions, how, and how to counter it	23
Table 7. Key experiments and studies at a glance	24
Table 8. Documented test cases and the logical check that exposes each	26
Table 9. Search test cases from AOFIRS Lesson 31	26
Table 10. A simple verification scale for research notes	28
Table 11. Verification methods that need no specialist software	28
Table 12. Types of bias, how to identify them, and how to counter them	30
Table 13. Quick-reference verification checklist	33
Table 14. Query templates for common verification tasks	37

Executive Summary

Many of us grew up believing that if we saw something, it must have really happened. Today, that is no longer true. Technology can create realistic faces in seconds, software can run entire news sites, and sponsored posts can look just like independent reviews. This report looks at a practical question for anyone researching online: how can someone figure out what is real, what is advertising, and what is biased, using logic and simple tools instead of advanced **software**?

To answer this, we combined material from AOFIRS Lesson 31 on searching news, articles, and blogs ([AOFIRS, n.d.](#)) with results from peer-reviewed studies, large surveys, regulatory reports, and real-world examples. Every statistic in this report comes from a named source in the APA reference list, and each citation links directly to its entry.

The one-sentence takeaway

Do not ask "does this look real?" Ask "who made this, where did it first appear, who else confirms it, and who benefits if I believe it?"

Key Findings

- **False stories move faster than true ones.** In a study of roughly 126,000 rumor cascades on Twitter between 2006 and 2017, falsehoods were 70% more likely to be retweeted than the truth ([Vosoughi et al., 2018](#)).
- **People are worried, and trust is falling.** In 2026, 62% of people across 48 markets said they were concerned about telling real from fake online, up four points in a year, while overall trust in news fell to 37%, the lowest level recorded since 2015 ([Reuters Institute for the Study of Journalism, 2026](#)).
- **Our eyes are poor deepfake detectors.** A meta-analysis of 56 papers and 86,155 participants put human accuracy on deepfake items at 55.5%, with confidence intervals that cross chance ([Diel et al., 2024](#)). People also overrate their own skill ([Köbis et al., 2021](#)).
- **Hidden advertising mostly goes unnoticed.** In controlled experiments, only 7% to 18% of readers recognized native advertising as paid content ([Wojdyski & Evans, 2016](#)), and 82% of middle-school students failed to identify a clearly labelled "sponsored content" item as an ad ([McGrew et al., 2017](#)).
- **Disclosure is the exception, not the rule.** An EU sweep of 576 influencers found that 97% posted commercial content, but only 20% consistently said so ([European Commission, 2024](#)).
- **Skills can be taught quickly.** After a short online course, the number of students who checked a website by reading what others said rose from 3 of 87 to 67 of 87 ([Breakstone et al., 2021](#)). A

15-minute game lowered the perceived reliability of fake posts without lowering trust in real news ([Roozenbeek & van der Linden, 2019](#)).

- **Logic beats eyesight.** Because surface clues are unreliable, the strongest defences against deepfakes are context checks, provenance tracing, and confirmation through a separate, trusted channel. This step could have prevented the \$25 million Arup video-call fraud ([Magramo, 2024](#)).

1. Introduction

1.1 Why This Report Matters

The internet has made research easier and harder at the same time. Easier, because almost any document, archive or eyewitness video is a few clicks away. Harder, because the same channels carry fabricated stories, synthetic media, paid persuasion that looks like journalism, and coordinated campaigns designed to look like grassroots opinion. For an internet research specialist, the cost of getting this wrong is real: a single unverified claim can travel from a draft into a client report, a classroom, or a published article, and from there into other people's work.

This report is written for researchers, educators, students, journalists and curious members of the public. It favors methods that need nothing more than a browser, a search engine and a habit of asking good questions. Where specialist tools exist, we mention them only as optional extras.

1.2 Scope and Research Questions

The report addresses six linked questions:

1. What methods reliably detect fake or misleading news?
2. How can deepfakes be deciphered through logical verification rather than complicated software?
3. How can advertising that hides inside articles, blogs and social posts be told apart from independent content, and how can biased news be recognized?
4. How do social media murmurs, advertisers, marketing companies and lobbying groups create false impressions?
5. What do experiments and documented test cases show, and what do their statistics actually mean?
6. What makes verification difficult, which biases affect both sources and researchers, and how can they be identified and countered?

1.3 How the Evidence Was Gathered

We started from the uploaded AOFIRS Lesson 31, which covers news-mode searching, advanced search operators, URL structures, time-date stamps, datelines, web archives, regional and language effects, newswire timestamp protocols, and blog searching ([AOFIRS, n.d.](#)). We then searched for primary studies and authoritative reports on each research question, read the original papers or official pages wherever they were accessible, and recorded exact figures, sample sizes, and caveats.

We gave different importance to different types of sources. Peer-reviewed studies and meta-analyses are the main sources for claims about human behavior. Large surveys, like those from the Reuters Institute Digital News Report and Pew Research Centre, are used for public attitudes. We use regulators such as the U.S. Federal Trade Commission and the European Commission for rules and enforcement.

Trusted news outlets help document specific events. We use commercial sources only when needed and clearly note when the publisher has something to sell, like the iProov deepfake survey ([iProov, 2025](#)) and NewsGuard trackers ([NewsGuard, 2026](#)).

A note on source bias in this report

Much of the experimental literature comes from the United States and Western Europe, often uses online convenience samples, and tests English-language material. Results may not transfer evenly to other countries, languages or age groups. Where a figure comes from a self-selected sample or an industry publisher, the text says so.

2. Key Concepts and Definitions

Clear words make clear thinking. "Fake news" has become a slogan that people use to dismiss anything they dislike, so researchers need more precise terms.

2.1 Misinformation, Disinformation and Malinformation

Clear words make clear thinking. Using clear language helps us think clearly. The term "fake news" is often used as a catch-all to dismiss things people do not like, so researchers need to use more precise words.

[Wardle & Derakhshan \(2017\)](#) proposed a framework for the Council of Europe that separates false content by intent and harm. It remains the most widely used vocabulary in the field. [Lazer et al. \(2018\)](#) add a useful working definition of fake news as fabricated information that mimics news media content in form but not in organisational process or intent, and [Allcott & Gentzkow \(2017\)](#) focus on articles that are intentionally and verifiably false and could mislead readers.

Table 1. Three types of information disorder

Type	What it is	Everyday example
Misinformation	False or misleading content shared without intent to harm. The sharer usually believes it.	A relative forwards an old flood photo as if it were today's storm.
Disinformation	False content created or shared knowingly to cause harm or gain advantage.	A fabricated quote attributed to a candidate, pushed by coordinated accounts before a vote.
Malinformation	Genuine information shared to cause harm, often by moving it out of its private or original context.	Leaking a real private email selectively to embarrass someone.

Note. Adapted from Wardle & Derakhshan (2017).

2.2 Deepfakes and Cheap Fakes

A **deepfake** is audio, video, or imagery generated or altered by machine learning so that a person appears to say or do something they never did. [Paris & Donovan \(2019\)](#) helpfully place deepfakes on a spectrum with **cheap fakes**: manipulations made with ordinary editing, such as slowing a clip, cropping out context, mislabeling a real photo, or splicing audio. In practice, cheap fakes remain extremely common, and the verification methods in this report work on both.

2.3 Hidden Advertising

Native advertising is paid content designed to match the look and feel of the surrounding editorial content. It includes "sponsored content," "partner content," advertorials, branded articles, paid influencer posts, affiliate-driven "best of" lists, and press releases that are republished as if they were

news. The U.S. Federal Trade Commission treats an ad's format as deceptive if it materially misleads consumers about the content's commercial nature ([Federal Trade Commission, 2015](#)).

2.4 Astroturfing, Front Groups and Manufactured Doubt

Astroturfing is hidden, coordinated campaigning that imitates spontaneous grassroots opinion. **Front groups** are organizations with neutral or public-interest names that advance the interests of undisclosed funders. **Manufactured doubt** is a strategy in which an industry funds or promotes research and messaging that make settled evidence look uncertain. Each is examined in Section 7.

3. Why False Content Travels So Well

Before we talk about how to spot false content, it helps to know why it spreads so easily. Research shows that it is mostly due to how people pay attention and how platforms are designed, not because people are foolish.

3.1 Novelty and Emotion

[Vosoughi et al. \(2018\)](#) studied about 126,000 contested news cascades tweeted by about 3 million people, totaling more than 4.5 million tweets. False news travelled farther, faster, deeper, and more broadly than the truth in every category, and the effect was strongest for political news. Falsehoods were 70% more likely to be retweeted, even after controlling for the original account's age, activity, follower counts, and verified status. The authors linked this to novelty: false stories tend to be more surprising and provoke stronger emotions, such as surprise and disgust ([Vosoughi et al., 2018](#)).

3.2 Inattention Rather Than Ideology

It is tempting to assume that people believe false headlines because they want to. [Pennycook & Rand \(2019\)](#) found something more hopeful: susceptibility to fake news was better explained by a lack of careful reasoning than by partisan motivated reasoning. People who scored higher on reflective thinking were better at telling true from false headlines regardless of whether those headlines suited their politics. Later experiments showed that simply nudging people to think about accuracy improved the quality of what they shared ([Pennycook et al., 2021](#)). This aligns with the familiar distinction between fast, intuitive thinking and slower, deliberate thinking ([Kahneman, 2011](#)).

3.3 Repetition Breeds Belief

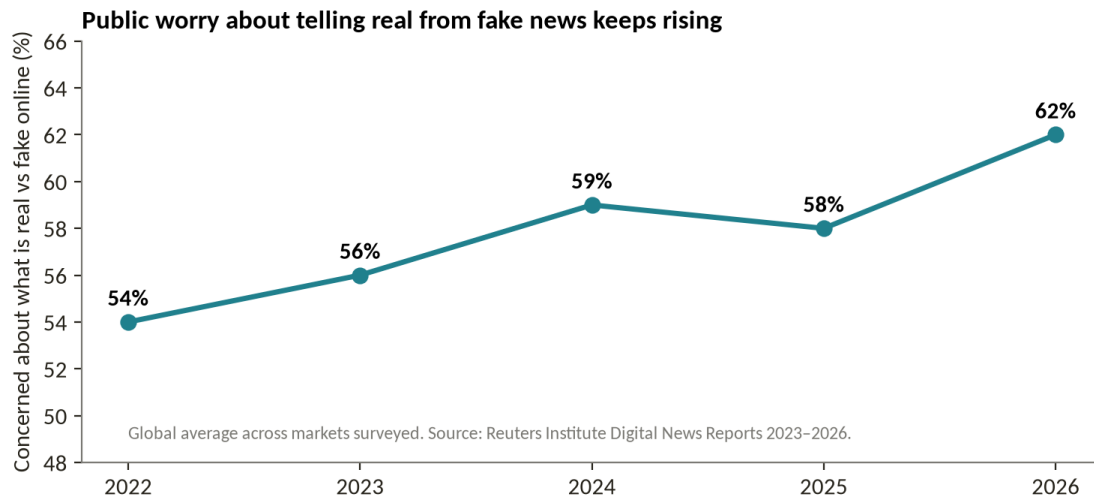
Researchers first documented the illusory truth effect in 1977, when repeated statements were rated as more likely to be true than new ones ([Hasher et al., 1977](#)). [Pennycook et al. \(2018\)](#) showed the same applies to fake news headlines: a single prior exposure increased their perceived accuracy, and the effect persisted a week later. For researchers, this means that seeing a claim in many places is not, by itself, evidence. Ten copies of one press release are still one source.

3.4 Who Shares, and Where People Now Get News

Sharing of fake news is concentrated. In a linked survey and Facebook data from the 2016 U.S. election period, 91.5% of respondents shared no links from fake-news domains, and users over 65 shared nearly seven times as many such articles as the youngest group ([Guess et al., 2019](#)). At the same time, where people get news is shifting. In 2026, social media and video networks became the leading online news source globally (54%), ahead of news organisations' own sites and apps (51%) ([Reuters Institute for the Study of Journalism, 2026](#)). In the United States, about one in five adults, and 37% of those under 30, regularly get news from "news influencers" on social media ([Stocking et al., 2024](#)). Trust is lowest exactly

where use is growing: only 22% trust news they see on social media and 20% trust news from AI chatbots ([Edelman, 2026](#)).

Figure 1. Share of people concerned about what is real and fake online, 2022–2026

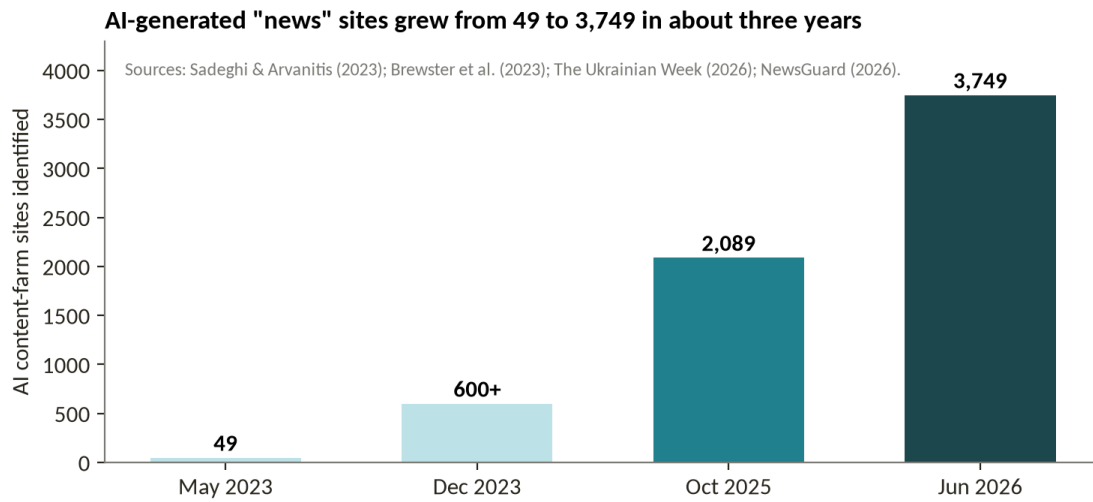


Note. Global averages across markets surveyed. The 2022 value is derived from the 2023 report's statement of a two-point rise. Data from Reuters Institute for the Study of Journalism (Reuters Institute for the Study of Journalism, 2023; Reuters Institute for the Study of Journalism, 2024; Reuters Institute for the Study of Journalism, 2025; Reuters Institute for the Study of Journalism, 2026).

3.5 The Machine-Made Content Surge

Generative AI has lowered the cost of producing plausible text to almost zero. NewsGuard identified 49 websites that appeared to be written almost entirely by AI in spring 2023 ([Sadeghi & Arvanitis, 2023](#)), more than 600 by December 2023 ([Brewster et al., 2023](#)), 2,089 by autumn 2025 as reported by [The Ukrainian Week \(2026\)](#), and 3,749 across 16 languages by June 2026 ([NewsGuard, 2026](#)). These "content farms" often carry generic names, publish hundreds of articles a day, and earn money from programmatic advertising. NewsGuard's audits also found that leading AI chatbots repeated false claims on news topics more than 28% of the time in January 2026 ([NewsGuard, 2026](#)), reminding us that an AI answer is a starting point for verification, not a substitute.

Figure 2. Growth of AI-generated "content farm" news sites tracked by NewsGuard



Note. Sources: (Sadeghi & Arvanitis, 2023; Brewster et al., 2023; The Ukrainian Week, 2026; NewsGuard, 2026). NewsGuard is a commercial rating company; counts reflect its own identification criteria. The December 2023 value was reported as "more than 600."

4. Detecting Fake News: Methods That Work

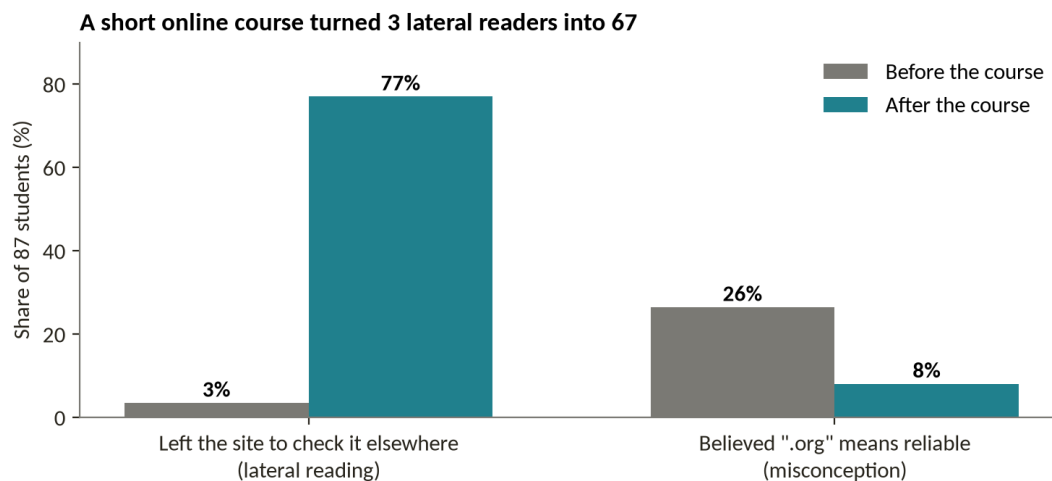
The good news is that the best ways to spot fake news are simple habits, not special software. Research shows that the techniques used by professional fact-checkers can be taught to students and the public in just a few hours.

4.1 Read Laterally, Not Vertically

[Wineburg & McGrew \(2019\)](#) watched 10 PhD historians, 10 professional fact-checkers and 25 Stanford undergraduates evaluate live websites on topics such as bullying, the minimum wage and teacher tenure. The historians and students mostly read **vertically**: they stayed on the page and studied its design, its "About" section, and its references. The fact-checkers read **laterally**: after a quick scan, they left the site, opened new tabs, and searched for what other, independent sources said about the organization behind it. The fact-checkers reached better-supported conclusions in a fraction of the time.

This matters because a polished site, a ".org" address, and a list of references are all easy to fake. What is hard to fake is an organization's reputation as described by others. When [Breakstone et al. \(2021\)](#) taught lateral reading in a short online course, the share of 87 college students who left a target website to check it rose from 3 to 67, average scores rose from 3.95 to 7.08 out of 13, and the number who believed a ".org" domain signaled reliability fell from 23 to 7. A district-wide field study in high school government classes later reported gains from a six-lesson lateral reading curriculum ([Wineburg et al., 2022](#)).

Figure 3. Lateral reading before and after a short online course (n = 87)



Note. Percentages calculated from counts reported by Breakstone et al. (2021).

4.2 The SIFT Moves

[Caulfield \(2019\)](#) condensed the fact-checkers' habits into four moves, known as SIFT. They are easy to remember and work on any content:

- **Stop.** Notice your emotional reaction and ask whether you know the source. If you are about to share, pause first.
- **Investigate the source.** Spend a minute finding out who is behind the content and what their expertise and agenda are, using other sites rather than the source's own claims.
- **Find better coverage.** Look for trusted reporting on the same claim. If the story is real and significant, others will have it.
- **Trace claims, quotes, and media to the original context.** Follow the chain back to the first publication, the full video, or the original study.

4.3 Why Checklists Alone Fall Short

Many people were taught to judge a site by checklist: Is there an author? Is it current? Does it look professional? Does it have a ".org" or ".edu" address? The Stanford History Education Group found that these surface checks fail badly. In one task, students evaluated MinimumWage.com, a site run by the Employment Policies Institute. Only 9% of 95 Advanced Placement history students and 6% of 58 college students identified the real backers behind it, even though a simple search for the organisation's name plus "funding" surfaces exposés about its ties to a Washington, D.C., lobbyist ([McGrew et al., 2017](#); [Stanford History Education Group, 2016](#)). A checklist looks inward at the page; good verification looks outward at the web.

4.4 Using Search Operators to Find the Original Story

The [CIRS Training Guide](#) Lesson 31 shows that vague searches often fail to retrieve the precise story a researcher needs, and that adding structure fixes this ([AOFIRS, n.d.](#)). A single-keyword query on one domain, such as "deepfake" restricted to ft.com, returns several stories from different years. Adding a date range narrows the results, and adding the exact title, URL fragments, and dateline text pinpoints the original article. Google documents the "before:" and "after:" date operators in its own search help pages ([Google, n.d.](#)).

The lesson identifies four pieces of information that, along with the domain, make a strong query string: the site's link structure, the article title, the publication date, and the author's name. It also explains that URLs come in three forms: breadcrumb paths, semantically readable paths, and encoded paths with parameters such as a document ID. The "inurl:" operator works well on readable paths but fails on encoded ones, as with Financial Times content URLs made up of a long alphanumeric string ([AOFIRS, n.d.](#)).

Table 2. Search-operator patterns for tracing a news story to its source

Goal	Example query pattern	Why it helps verification
Find all coverage on one outlet	site:ft.com "deepfake"	Shows whether the outlet actually published the claimed story.

Goal	Example query pattern	Why it helps verification
Narrow to a time window	site:ft.com "deepfake" after:2020-12-31 before:2021-03-31	Separates the original from later rewrites and recycled versions.
Pinpoint an exact article	site:edition.cnn.com inurl:business inurl:2019/02/01 intitle:"what is a deepfake"	Uses the site's own URL structure to land on the original page.
Use the right date format	site:forbes.com inurl:2021/03/17 "google algorithms" intext:"david hunter"	Lesson 31 shows that only the URL date format worked; text date formats failed.
Find blogs on a topic	internet "search engine" semantics inurl:blog inurl:blogs	Separates personal or company blogs from news reporting.
Exclude a dominant site	... -site:researchgate.com	Removes a site that crowds out independent sources.

Note. Patterns adapted from AOFIRS (n.d.). Operator behavior changes over time; test any query before relying on it.

4.5 Reading the Clock: Timestamps, Datelines and Archives

Time is one of the most abused elements in false content. Lesson 31 warns that "timestamps are a mess on the internet," and that a common false claim is that a page or video is dated before an event, which is then treated as proof the event was staged ([AOFIRS, n.d.](#)). Researchers need to distinguish several kinds of date:

Table 3. Kinds of date a researcher may encounter

Date type	What it represents	How to use it
Publication date	When the page was first uploaded by the publisher.	The primary date for citation; may be absent, removed or changed.
Last-updated stamp	When the story was last revised; many outlets show both first-published and updated times.	Essential for developing stories; check what changed.
Dateline	Where (and often when) the story was filed, not necessarily where the event happened.	Lesson 31's New York Times example was filed from Cologne about a crash in Bavaria.
Event time	When the incident actually occurred, often stated inside the text.	Compare with the publication date; impossible gaps are a red flag.
Indexing date	When a search engine first discovered the page.	A rough approximation of creation; not reliable on its own.
Archive capture date	When an archive such as the Wayback Machine saved a copy.	Proves a page existed in a given form on that date.
Newswire timestamp	The time a wire story was issued; rules differ by agency.	Lesson 31 notes that some wires keep the first-issue time, while others update it with revisions.

Note. Compiled from AOFIRS (n.d.), which draws on Associated Press style guidance and Factiva newswire documentation.

Two updates to the lesson are worth noting. First, Google removed cached-page links from its results in early 2024 and announced it would also retire the "cache:" operator ([Porter, 2024](#)), so researchers should rely on public archives instead. Second, WebCite, listed in the lesson's archive table, no longer accepts new archiving requests, although older snapshots remain viewable ([WebCite, n.d.](#)). The Internet Archive's Wayback Machine remains the main resource; in October 2025 it passed one trillion archived web pages ([Adams, 2025](#)). Lesson 31's worked example of recovering pages from the retired DMOZ directory through the Wayback Machine is still an excellent model.

4.6 Escaping the Local Bubble: Region and Language

Search engines localise results by IP address, account settings, and query language. Lesson 31 demonstrates that changing the "hl" (language) and "gl" (country) parameters in a Google News URL shifts the edition, and that typing the same query in Ukrainian rather than English surfaces Ukrainian news agencies instead of British and American outlets. Importantly, the lesson notes that changing the browser's interface language does not change results; changing the query language does ([AOFIRS, n.d.](#)). For verification, this powerful, free technique works. If a dramatic claim about another country appears only in your country's outlets and nowhere in that country's press, dig deeper.

4.7 Transparency Signals and Newswires

Google looks for transparency signals when evaluating news sources, including clear dates, bylines, author bios, contact information, and publisher background ([Granito, 2021](#); [Southern, 2021](#)). Researchers can use the same list as a quick screen. Missing signals don't prove falsehood, since some regions make bylines dangerous for journalists, but they shift the burden of proof.

Lesson 31's newswire table includes PR Newswire and Business Wire alongside the Associated Press, Reuters, and Agence France-Presse ([AOFIRS, n.d.](#)). This is an important distinction. The first two distribute press releases that organisations write and pay for; the others are news agencies that employ journalists. A news site can republish a press release so it looks like reporting. Checking whether a story originated on a press-release wire is one of the quickest ways to identify promotional content dressed as news.

5. Deciphering Deepfakes Without Complicated Software

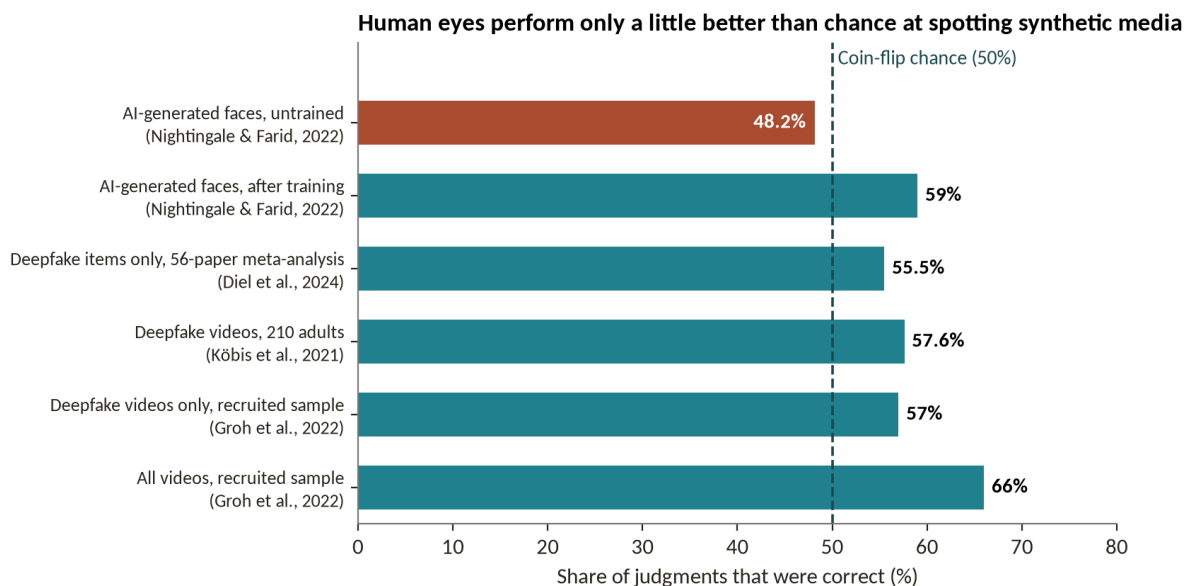
5.1 What the Experiments Tell Us

The experimental record is humbling. In two experiments, 315 participants classifying faces as real or AI-generated scored 48.2%, slightly below a coin flip; even with training and feedback, accuracy rose only to 59.0%. Worse, participants rated synthetic faces 7.7% more trustworthy than real ones ([Nightingale & Farid, 2022](#)). In a study of 210 adults viewing short videos, accuracy was 57.6%; neither warning people about deepfakes nor paying them for correct answers improved it, and participants were biased toward calling deepfakes authentic while overestimating their own ability ([Köbis et al., 2021](#)).

The largest synthesis to date, a meta-analysis of 56 papers and 86,155 participants, found that accuracy on deepfake stimuli was 55.5% overall: 62.1% for audio, 57.3% for video, 53.2% for images and 52.0% for text, with confidence intervals crossing the 50% chance line ([Diel et al., 2024](#)). An industry survey of 2,000 UK and U.S. consumers reported that only 0.1% correctly classified every real and fake item shown, even when told to look for fakes, and that participants were 36% less likely to identify a synthetic video than a synthetic image ([iProov, 2025](#)). Because the publisher sells identity-verification technology, that survey should be read as illustrative rather than definitive.

There is one encouraging nuance. In a study of 15,016 participants, [Groh et al. \(2022\)](#) found that ordinary people were roughly as accurate as the leading computer model overall, that crowd judgment was often more accurate than individuals, and that people did better than the model on videos of well-known political leaders. Context and familiarity with the person helped. However, when the model gave wrong advice, people often followed it, a warning about over-trusting automated detectors.

Figure 4. Human accuracy at identifying synthetic media across major studies



Note. Values are mean accuracies reported by each study. Diel et al. (2024) value is for deepfake stimuli only. Groh et al. (2022) values are for the recruited sample in Experiment 2.

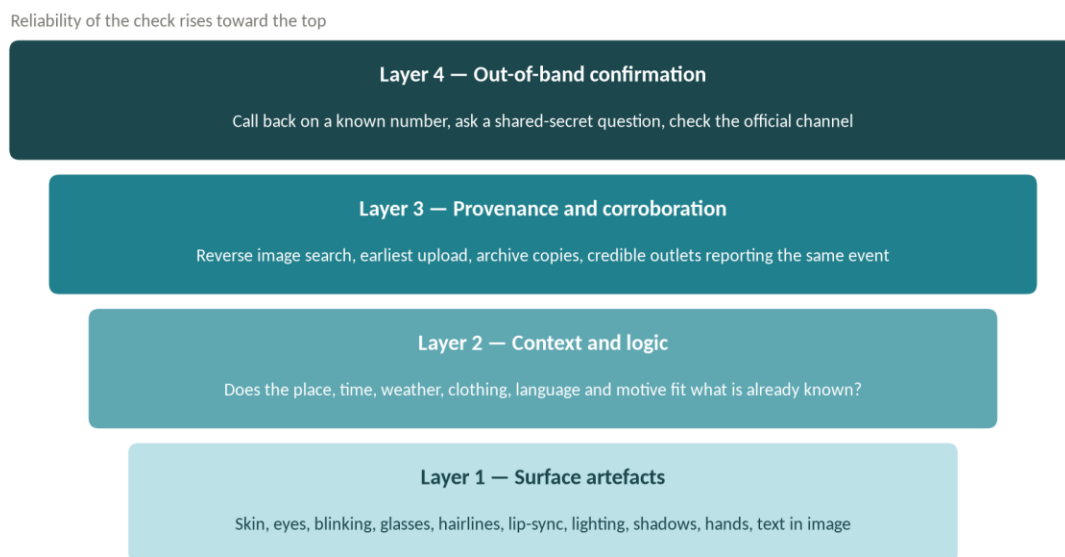
What this means in practice

If trained observers barely beat chance by looking, the answer is not to look harder. It is to switch from "Does this look fake?" to "Can this be confirmed?" Logical verification shifts the burden from your eyes to the evidence.

5.2 A Layered Approach

We suggest thinking about deepfake verification in four layers, from the least to the most reliable. Surface clues can be helpful, but your final judgment should rely on the higher layers.

Figure 5. Four layers of logical deepfake verification



Note. This report's framework is developed from published guidance ([MIT Media Lab, n.d.](#); [Farid, n.d.](#); [Caulfield, 2019](#); [Silverman, 2014](#)).

5.3 Layer 1: Surface Artifacts

MIT Media Lab's Detect Fakes project suggests eight areas to examine: the face, the cheeks and forehead, the eyes and eyebrows, glasses, facial hair, facial moles, blinking, and lip movements ([MIT Media Lab, n.d.](#)). Hany Farid, a digital forensics researcher, adds physical-consistency checks anyone can do by eye: light and shadows should align with a single dominant light source, reflections should align geometrically, and lines in a scene should converge consistently ([Farid, n.d.](#)).

Table 4. Surface checks for images, video and audio

Area	What to look for	Caution
Skin and face	Skin too smooth or too wrinkled for the person's age; face edges that blur or shimmer against hair or background.	Beauty filters and compression produce similar effects.
Eyes and blinking	Unnatural blinking rate, mismatched reflections in the two eyes, eyebrows that do not move with the face.	Newer models have largely fixed early blinking flaws.
Glasses and jewelry	Glare that does not shift as the head moves; earrings that differ between ears.	Real lighting can also be odd.
Mouth and speech	Lip shapes that do not match sounds such as "m," "b" and "p"; teeth that blur into one block.	Poor video connections cause lip-sync lag.
Light and shadows	Shadows pointing in different directions under one sun; reflections that do not line up.	Requires careful observation of the whole scene.
Hands, text and background	Wrong number of fingers, warped signage, gibberish text, objects that melt into each other.	A clean result proves nothing; these flaws are disappearing.
Audio	Flat emotion, odd breathing, unnatural pauses, background noise that cuts abruptly.	Audio deepfakes are the hardest to spot by ear.

Note. Compiled from sources (MIT Media Lab, n.d.; Farid, n.d.). Absence of artifacts is never proof of authenticity.

5.4 Layer 2: Context and Logic

This is where most fakes fall apart. Ask the questions a detective would ask:

- **Is it plausible?** Would this person say this, in this way, in this setting? Does it contradict their documented record?
- **Does the setting fit?** Check the weather, season, time of day, landmarks, signage language, license plates, uniforms, and clothing against what is known about the claimed place and date.
- **Who released it first, and why now?** Deepfakes often appear at moments of maximum impact: just before an election, during a crisis, or as markets open.
- **What is the call to action?** Urgent requests to transfer money, stay home from voting, surrender, or panic are classic signs of manipulation.
- **Is anything missing?** A dramatic event with only one video, from one account, with no other witnesses is a serious warning sign.

5.5 Layer 3: Provenance and Corroboration

Provenance means the documented history of a piece of content. Without specialist software, a researcher can still do a great deal:

- **Reverse image search.** Upload an image or a screenshot of a video frame to a mainstream reverse image search such as Google Lens, Bing Visual Search, or TinEye to find earlier copies. An "About this image" style panel can show when an image was first indexed.
- **Find the earliest upload.** Sort results by date and look for the first account to post it. A "breaking" image that appeared years ago is a recycled cheap fake.
- **Check the archives.** Use the Wayback Machine to see whether a page, profile, or statement existed in the claimed form on the claimed date, as Lesson 31 demonstrates ([AOFIRS, n.d.](#)).
- **Look for independent corroboration.** Look for credible outlets in multiple countries and languages, official channels of the person or agency involved, and established fact-checkers.
- **Look for Content Credentials.** Some cameras, editing programs, and platforms now attach provenance data under the C2PA standard, which records facts about the history of an image, video, or document ([Coalition for Content Provenance and Authenticity, n.d.](#)). Its presence helps; its absence proves nothing, because most content still lacks it.

5.6 Layer 4: Out-of-Band Confirmation

The most reliable check is to confirm through a channel the fraudster doesn't control. In early 2024, an employee in Arup's Hong Kong office joined a video call with people who looked and sounded like the company's chief financial officer and colleagues, and then made 15 transfers totalling about US\$25.6 million to fraudsters ([Magramo, 2024](#)). A call back to the executive on a known phone number, or a pre-agreed verification question, would have broken the illusion. The same logic applies to voice-cloned calls from "relatives" in distress and to fake executive voice messages.

5.7 The Liar's Dividend

Deepfakes create a second problem: they let people dismiss genuine evidence as fake. Researchers should apply the same standards in both directions. A claim that a real recording is a deepfake also needs evidence, and it should follow the same provenance and corroboration steps.

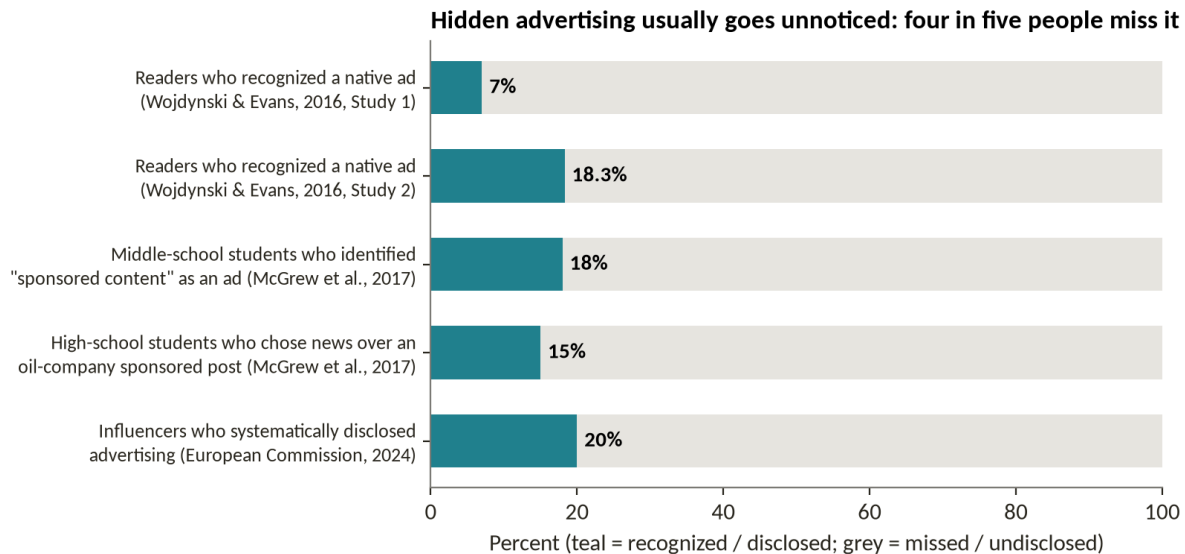
6. Spotting Advertising Hidden in Content

6.1 The Scale of the Problem

Paid content works partly because readers do not notice it. In experiments by [Wojdyski & Evans \(2016\)](#), only 7% of participants in the first study and 18.3% in the second recognised a native article as paid advertising; disclosures placed in the middle or at the bottom of the article, and labels using the words "advertising" or "sponsored," worked better than other options. A representative survey experiment of 800 U.S. adults found that fewer than one in ten recognised native advertising. However, well-designed disclosure labels improved recognition ([Amazeen & Wojdyski, 2017](#)); in another experiment with 183 adults, 73% who saw an in-feed ad without a disclosure failed to recognise it as advertising before clicking ([An et al., 2019](#)).

Young people struggle too. In Stanford's assessments, 82% of middle-school students failed to identify a clearly labelled "sponsored content" item as an advertisement. Nearly 70% of more than 200 high school students judged an oil company's sponsored post more reliable than a traditional news story, with only 15% choosing the news article ([McGrew et al., 2017](#)). In a later national sample of 3,446 high school students, two-thirds could not tell news stories from ads marked "Sponsored Content" on a popular site's homepage ([Breakstone et al., 2019](#)).

Figure 6. Recognition and disclosure of hidden advertising across studies



Note. Middle-school value is the complement of the 82% who failed to identify sponsored content. Sources: ([Wojdyski & Evans, 2016](#); [McGrew et al., 2017](#); [European Commission, 2024](#)).

6.2 Common Formats

Table 5. Formats of hidden advertising and how to recognize them

Format	How it appears	Tell-tale signs
Native or sponsored article	An article in a news site's layout, written or paid for by a brand.	Small labels such as "Sponsored," "Partner content," "Paid post," "Brand studio"; a brand named repeatedly and favorably; no critical voices.
Advertorial	Long-form promotional copy styled as editorial, often in magazines or health sites.	A product solution introduced after a "problem" story; testimonials; purchase links.
Influencer post	A personal-seeming photo, video or story recommending a product.	"#ad," "#sponsored," "Paid partnership" toggles, discount codes, affiliate links, gifted-product mentions.
Affiliate "best of" review	Rankings of products on blogs or review sites.	Every product links to a retailer; disclosure buried at the bottom; no testing method described.
Press release as news	A wire release republished by news sites with little change.	Identical wording across many sites; "PRNewswire" or "Business Wire" credit; quotes only from the company.
Sponsored research or white paper	A "study" or survey released by a company with commercial interest in the result.	Funder named in fine print; methods not published; findings that conveniently support the funder's product.
Fake or incentivized reviews	Star ratings and testimonials on shopping and map sites.	Bursts of five-star reviews in a short period, generic wording, reviewers with no history.

Note. Compiled from the [Federal Trade Commission \(2015, 2019, 2024\)](#) and the [European Commission \(2024\)](#), and from the studies cited in Section 6.1.

6.3 A Quick Test for Hidden Persuasion

Ask five questions of any article, blog post, or video that seems to recommend something:

1. **Is there a label?** Look above the headline, beside the byline, at the very bottom, and in the caption. Search the page for "sponsored," "partner," "paid," "affiliate," and "advertisement."
2. **Who wrote it?** Is the author a staff journalist, a contributor, a "brand studio," or a company employee? Read laterally about them.
3. **Is there a product, brand, or call to action?** Purchase links, discount codes, and "learn more" buttons point to commercial intent.
4. **Are other perspectives present?** Independent journalism normally includes critics, competitors, or caveats. Promotional content rarely does.

5. **Where else does this exact text appear?** Paste a distinctive sentence in quotation marks into a search engine. Many identical copies suggest a press release or syndicated advertorial.

6.4 What the Rules Require

In the United States, the FTC's 2015 enforcement policy states that an ad is deceptive if its format misleads consumers about its commercial nature ([Federal Trade Commission, 2015](#)). Its guidance for influencers says that any material connection to a brand, whether payment, free products, or a family relationship, must be disclosed clearly and obviously ([Federal Trade Commission, 2019](#)). In August 2024, the FTC announced a final rule banning fake reviews and testimonials, including buying reviews and AI-generated fake reviews ([Federal Trade Commission, 2024](#)). In the European Union, the 2024 influencer sweep found that 38% of influencers did not use the platforms' own commercial-content labels and 30% gave no company details at all ([European Commission, 2024](#)). Rules exist; compliance is patchy. Researchers should assume that disclosure may be missing and check for commercial ties themselves.

7. How Murmurs, Marketers and Lobbyists Distort Impressions

Not all distortion comes from outright lies. Much of it comes from **creating a false impression of consensus or doubt**: making a fringe view look popular, or making a settled finding look contested. Understanding the playbook helps researchers recognize it.

7.1 Social Media Murmurs and Coordinated Campaigns

A "murmur" is the background buzz of posts, comments, and shares that makes an idea feel widely held. It can arise naturally, but it can also be manufactured. Studying accounts identified in South Korean court proceedings, [Keller et al. \(2020\)](#) showed how a state intelligence agency ran a hidden campaign during the 2012 presidential election, and that the accounts gave themselves away through coordination patterns: posting the same content at the same time and retweeting each other far more than genuine users do. In 2017, Facebook told a Senate committee that content produced by Russia's Internet Research Agency may have reached up to 126 million users ([PBS NewsHour, 2017](#)).

7.2 Front Groups and Think-Tank Branding

The MinimumWage.com example in Section 4.3 is a classic front-group case: a professional-looking site with a neutral name and a ".com" address, run by an organisation with undisclosed ties to industry interests ([Stanford History Education Group, 2016](#)). Front groups often use words such as "Institute," "Alliance," "Coalition," "Council," "Citizens for," or "Centre for" in their names. None of these words is suspicious in itself; the test is whether funding and leadership are transparent and whether independent sources describe the group the same way it describes itself.

7.3 Manufactured Doubt

In 1969, a tobacco company memo stated plainly: "Doubt is our product since it is the best means of competing with the 'body of fact'" ([Brown & Williamson, 1969](#)). [Oreskes & Conway \(2010\)](#) documented how a small group of scientists and public relations campaigns used similar tactics on issues from tobacco smoke to climate change. Internal documents [Kearns et al. \(2016\)](#) analysed show that the Sugar Research Foundation sponsored a 1965 literature review in a leading medical journal and, through the 1960s and 1970s, a research program that cast doubt on sugar's hazards while promoting fat as the main dietary culprit in heart disease.

Funding bias is not only historical. A Cochrane review found that manufacturer-sponsored drug and device studies more often reported favourable efficacy results (risk ratio 1.27) and favourable conclusions (risk ratio 1.34) than studies with other funding ([Lundh et al., 2017](#)). Sponsorship does not make a study wrong, but it is a reason to look more closely at its design, outcomes, and independent replication.

7.4 News Influencers and Blurred Lines

Pew Research Centre identified 2,058 news influencers with at least 100,000 followers and found that about one in five U.S. adults regularly get news from them ([Stocking et al., 2024](#)). Influencers can be valuable, independent voices, but they are not bound by newsroom standards, and their income may depend on sponsorships, affiliate deals, or audience growth that rewards outrage. The same transparency questions apply: who pays them, and do they disclose it?

Table 6. Who distorts impressions, how, and how to counter it

Actor	Typical tactics	Warning signs	Researcher's counter
Coordinated social accounts	Mass posting, hashtag flooding, reply swarms, fake grassroots.	New accounts, identical wording, posts in synchronized bursts, profile photos that look synthetic.	Check account age and history; search for identical phrasing; look for coverage outside the platform.
Advertisers and brands	Native ads, advertorials, influencer deals, affiliate reviews.	Small labels, purchase links, discount codes, no critical voices.	Search for labels; read laterally on the author; find independent reviews.
Marketing and PR firms	Press releases as news, placed op-eds, sponsored surveys, review management.	Identical text across outlets, wire credits, company-only quotes.	Trace text to the wire; locate the original data and methods.
Lobbying groups and front groups	Neutral-sounding names, funded studies, "experts" with undisclosed ties.	Hidden funders, single-issue messaging, mismatch between self-description and outside descriptions.	Search the name plus "funding"; check lobbying disclosures and independent reporting.
State-linked operations	Fake outlets, impersonation, deepfakes, troll networks.	Spikes around elections or crises, foreign-language traces, impersonated logos.	Cross-check in multiple languages and regions; consult established fact-checkers.
AI content farms	Hundreds of auto-written articles a day, ad-funded.	Generic site names, no named staff, error messages left in text, recycled stories.	Check "About" and staff pages laterally; look for original reporting elsewhere.

Note. Synthesized from sources (Keller et al., 2020; Stanford History Education Group, 2016; Oreskes & Conway, 2010; Kearns et al., 2016; NewsGuard, 2026; Federal Trade Commission, 2019).

8. Experiments, Test Cases and Statistics

This section brings together experimental evidence, documented real-world cases, and search test cases worked through in [CIRS Training Guide](#) Lesson 31 of the CIRS Training Manual. It closes with a reflection on what the numbers can and cannot tell us.

8.1 Summary of Key Experiments

Table 7. Key experiments and studies at a glance

Study	Sample	Design	Key result
Vosoughi et al. (2018)	~126,000 cascades; ~3 million users	Observational analysis of Twitter, 2006–2017	Falsehoods 70% more likely to be retweeted than truth.
Guess et al. (2019)	1,191 with linked Facebook data	Survey linked to sharing behavior	8.5% shared any fake-news link; users over 65 shared about 7 times as many as the youngest.
Pennycook et al. (2018)	Multiple online experiments	Repeated exposure to headlines	One prior exposure raised perceived accuracy; effect persisted a week.
Wineburg & McGrew (2019)	10 historians, 10 fact-checkers, 25 students	Think-aloud website evaluation	Fact-checkers read laterally and reached better conclusions faster.
Breakstone et al. (2021)	87 college students	Pre/post online course	Lateral readers rose from 3 to 67; mean score 3.95 to 7.08 of 13.
Breakstone et al. (2019)	3,446 high school students	National assessment	52% took a Russian-filmed video as strong evidence of U.S. voter fraud; 96% missed a site's fossil-fuel ties.
Wojdyski & Evans (2016)	242 in Study 1	Experiments on disclosure position and wording	7% (Study 1) and 18.3% (Study 2) recognized native ads.
Nightingale & Farid (2022)	315 in Experiment 1	Classify real vs AI faces	48.2% accuracy; AI faces rated 7.7% more trustworthy.
Köbis et al. (2021)	210 adults	Classify 16 short videos	57.6% accuracy; incentives and warnings did not help; overconfidence.
Groh et al. (2022)	15,016 participants	Humans vs model vs combined	Humans about as accurate as leading model; crowds and context helped.
Diel et al. (2024)	56 papers; 86,155 people	Meta-analysis	Deepfake detection 55.5%, not reliably above chance.
Roozenbeek & van der Linden (2019)	~14,000 paired responses	Pre/post "Bad News" game	Perceived reliability of fake items fell from 2.61 to 2.06 (about 21%).
Roozenbeek et al. (2022)	6,464 (lab) + 22,632 (YouTube)	Prebunking videos	Better recognition of manipulation techniques across the political spectrum.
Pennycook & Rand (2022)	20 experiments; 26,863 people	Accuracy prompts	False-headline sharing down 10%; sharing discernment up 71.7%.

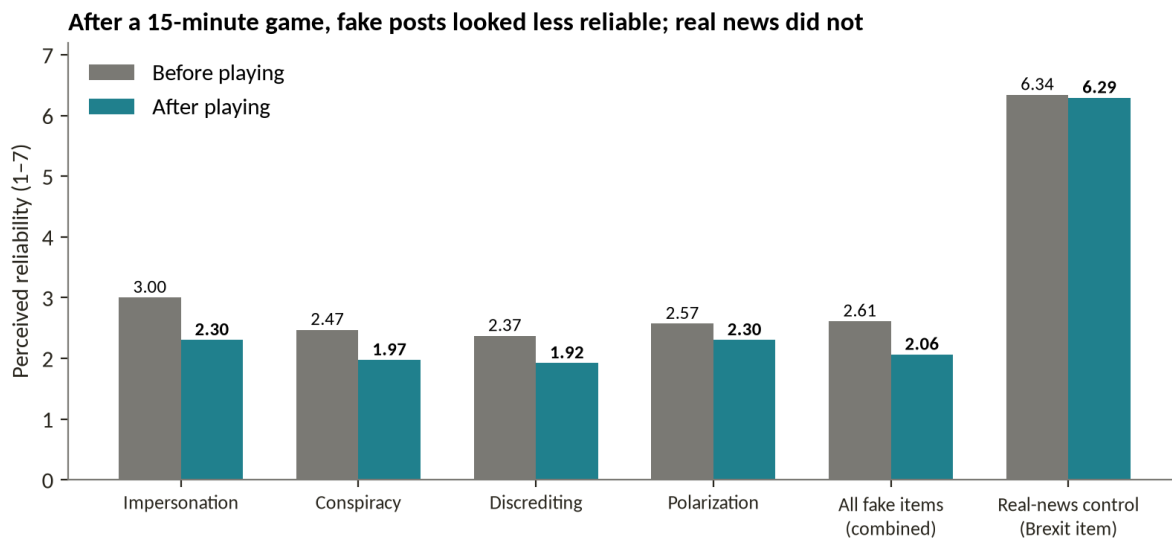
Study	Sample	Design	Key result
Walter et al. (2020)	30 studies; 20,963 people	Meta-analysis of fact-checking	Positive average effect ($d = 0.29$), weaker against strong prior beliefs.

Note. Synthesised from [Keller et al. \(2020\)](#); [Stanford History Education Group \(2016\)](#); [Oreskes & Conway \(2010\)](#); [Kearns et al. \(2016\)](#); [NewsGuard \(2026\)](#); and the [Federal Trade Commission \(2019\)](#).

8.2 What Interventions Achieve

Three intervention families have strong evidence. **Prebunking** (also called inoculation) exposes people to a weakened dose of manipulation techniques so they can recognise them later ([van der Linden, 2023](#)). In the "Bad News" game, players spent about 15 minutes acting as a fake-news producer. Afterwards, they rated fake posts using impersonation, conspiracy, discrediting, and polarisation techniques as less reliable, while their ratings of real-news control items barely changed ([Roozenbeek & van der Linden, 2019](#)). Across the combined fake items, perceived reliability dropped about 21%, consistent with the average reported in later coverage of the research ([Sleek, 2022](#)). Short prebunking videos tested in six randomised studies and a YouTube field study with 22,632 viewers produced similar gains ([Roozenbeek et al., 2022](#)).

Figure 7. Perceived reliability of posts before and after playing the "Bad News" game



Note. Ratings on a 1–7 scale, where 1 is unreliable. The sample was self-selected (75% male, 59% liberal), so results describe willing players rather than the general public—data from [Roozenbeek & van der Linden \(2019\)](#).

Accuracy prompts ask people to consider whether a headline is true before deciding what to share. Across 20 experiments with 26,863 participants, these prompts reduced intentions to share false headlines by 10% and improved sharing discernment by 71.7% relative to control ([Pennycook & Rand, 2022](#)). **Lateral reading instruction** produces some of the largest skill gains, as Section 4.1 shows ([Breakstone et al., 2021](#)). **Fact-checking** works on average, though modestly, with a meta-analytic effect size of $d = 0.29$ that weakens when corrections confront strongly held beliefs ([Walter et al., 2020](#)).

8.3 Real-World Test Cases

Table 8. Documented test cases and the logical check that exposes each

Case	What happened	Logical verification that works
Zelensky "surrender" deepfake (March 2022)	A deepfake video of Ukraine's president telling troops to lay down arms was posted after hackers breached a Ukrainian TV website; Meta removed it and Zelensky denied it (Holroyd & Olorunselu, 2022).	Plausibility (it contradicted his public stance); official channels (he publicly denied it); platform takedown confirmed manipulation.
AI image of Pentagon explosion (May 2023)	An apparently AI-generated image of an explosion near the Pentagon spread on Twitter and sparked a brief dip in the stock market (Bond, 2023).	Corroboration (no other witnesses, no local reporting); authorities quickly confirmed that no explosion had taken place.
Biden voice-clone robocall (January 2024)	Two days before the New Hampshire primary, calls with an AI-cloned voice of President Biden urged voters to stay home. The consultant behind it paid about \$500 to send them and was fined \$6 million (Shepardson, 2024).	Call to action (do not vote) is a classic red flag; check official campaign and election-office channels.
Arup video-call fraud (2024)	A Hong Kong employee transferred about US\$25.6 million in 15 transactions after a video call with deepfaked senior colleagues (Magramo, 2024).	Out-of-band confirmation: call back on a known number; require dual approval for unusual transfers.
Russian-filmed "ballot stuffing" video	In a national assessment, 52% of 3,446 students judged a grainy video "strong evidence" of U.S. voter fraud; it had been filmed in Russia. Only three students traced its source (Breakstone et al., 2019).	Trace to original context; a quick search finds articles exposing the video.
MinimumWage.com front group	A professional site on wage policy; only 6% of college students identified its backers (McGrew et al., 2017).	Lateral reading: search the organization name plus "funding."
Sugar industry research (1960s–1970s)	Industry-funded reviews downplayed sugar's role in heart disease (Kearns et al., 2016).	Follow the money: check funding statements and later independent evidence.

8.4 Search Test Cases from [CIRS Training Guide](#) Lesson 31

The CIRS Training Guide Lesson 31 includes several worked examples that function as test cases for search-based verification (AOFIRS, n.d.). They are summarized below, with the lesson's reported outcomes.

Table 9. Search test cases from AOFIRS Lesson 31

Test case	Procedure	Outcome and lesson
A. Locating an FT story on deepfakes	Query 1 used only the domain and keyword; Query 2 added a date range; Query 3 added inurl:, intitle: and intext: elements.	Query 1 returned stories from different years; Query 2 narrowed to the right period; Query 3 retrieved the exact article. Context beats keywords.
B. Locating a CNN explainer video	Built a query from the semantic URL: section, date path and title.	Matching the site's own URL structure reached the original 2019 page that was buried in the index.
C. Date-format test on Forbes	Three versions: text date "mar 17, 2021," ISO date "2021-03-17," and URL path "inurl:2021/03/17."	Only the URL-path version succeeded. Date formats must match how the site stores them.
D. Recovering a retired site (DMOZ)	Used inurl:dmoz.org to find surviving links, then retrieved archived pages from the Wayback Machine by date.	Archived copies preserved references from a directory that no longer exists.
E. OR syntax in blog searches	Compared [inurl:blog inurl:blogs] with [inurl:blog blogs].	About 32,200 vs 601,000 results: operator syntax changes the scope dramatically.
F. Language and region	Ran "gun control controversy" in English and in Ukrainian translation.	The Ukrainian query surfaced Ukrainian agencies; the English one surfaced UK and U.S. outlets. Query language steers coverage.

Note. Result counts are as reported in the lesson at the time of writing and will differ today.

8.5 Reflecting on the Statistics

Numbers are persuasive, so they deserve the same scrutiny as any other claim. Five cautions help keep these statistics in perspective:

- **Lab is not life.** Many detection studies ask people to judge short clips with a 50/50 chance of being fake. In real life, most content is genuine, people are not warned, and context is available. Accuracy could be higher or lower outside the lab.
- **Samples are often convenient.** The "Bad News" evaluation was self-selected and skewed toward young, male, liberal participants ([Roozenbeek & van der Linden, 2019](#)). Online panels can differ from the general population.
- **Effects can fade.** Inoculation and prompt effects are real but modest and may decay without reinforcement. A 10% reduction in false sharing is meaningful at scale but is not immunity.
- **The ground keeps moving.** Detection studies from 2021 used older generation methods. Fakes are becoming more realistic, so older accuracy figures may overstate what people can do today.
- **Different measures, different meanings.** Accuracy, sensitivity, sharing intentions and perceived reliability are not interchangeable. Always check what was measured before comparing figures across studies.

Taken together, the statistics tell a consistent story. People are poor at detecting manipulation by inspection. Still, they get much better when they are given simple, transferable strategies that move attention from the content itself to its source, origin, and corroboration.

9. The Problem of Verification

9.1 What "Verified" Should Mean

Verification is not a yes-or-no stamp. A careful researcher records a level of confidence and the reasons for it. We suggest four categories for working notes and reports:

Table 10. A simple verification scale for research notes

Status	Meaning	Minimum evidence
Verified	Confirmed by the original source and at least one independent, credible source.	Original document, data or recording located; independent corroboration.
Partly verified	Core facts confirmed but some details (numbers, dates, quotes) not yet checked.	Original source located; details flagged for follow-up.
Unverified	Plausible but not yet confirmed; must not be presented as fact.	Recorded with its source and a note that it is unconfirmed.
False or misleading	Contradicted by stronger evidence or shown to be fabricated, altered or out of context.	Documented contradiction, original context, or authoritative correction.

9.2 Verification Methods Compared

Table 11. Verification methods that need no specialist software

Method	Best for	Strength	Limitation
Lateral reading	Judging unfamiliar websites, organizations and authors.	Fast; strongly supported by experiments.	Depends on others having written about the source.
SIFT moves	Any claim, image or post on social media.	Memorable; covers source, coverage and origin.	Requires the discipline to stop.
Advanced search operators	Finding the original article, date and version.	Precise; free; covered in Lesson 31.	Operator behavior changes; some URL formats resist it.
Archive comparison	Proving what a page said and when.	Independent, timestamped record.	Not every page is archived.
Reverse image search	Finding earlier copies of photos and video frames.	Exposes recycled and out-of-context media.	Heavily edited or brand-new fakes may not match.
Cross-language and cross-region search	Checking claims about other countries.	Breaks the local filter bubble.	Translation can shift meaning, as Lesson 31 shows.
Out-of-band confirmation	Requests for money, credentials or urgent action.	Very hard for fraudsters to defeat.	Requires a known, trusted contact route.
Checklist tests (e.g., CRAAP)	Teaching basic awareness.	Simple to explain.	Inward-looking; easily fooled by polished sites.

Note. Assessments draw on sources (Wineburg & McGrew, 2019; Caulfield, 2019; Breakstone et al., 2021; Silverman, 2014; AOFIRS, n.d.).

9.3 Practical Obstacles

- **Missing or manipulated dates.** Lesson 31 notes that many publishers remove dates to recycle content, and self-publishers often omit them altogether ([AOFIRS, n.d.](#)).
- **Masked URLs.** Encoded and parameter-based URLs defeat operators like `inurl:`, so researchers need alternatives such as title and author searches.
- **Personalised and localised results.** Search history, location, and language shape what appears, so two researchers can see different evidence for the same query ([AOFIRS, n.d.](#)).
- **Disappearing content.** Pages are edited or deleted, cached copies are no longer offered by Google ([Porter, 2024](#)), and some archives no longer accept new captures ([WebCite, n.d.](#)).
- **Synthetic volume.** Thousands of AI content farms ([NewsGuard, 2026](#)) create the illusion that many sources repeat a claim, exploiting the illusory truth effect ([Pennycook et al., 2018](#)).
- **AI assistants that sound certain.** Chatbots can repeat false claims fluently ([NewsGuard, 2026](#)). Every claim, citation, and link they produce needs the same checks as any other source, or unverified material accumulates into what we might call verification debt.

9.4 Correcting Without Backfiring

Researchers often need to correct errors in their own work or in public. Early research raised fears that corrections could "backfire" and strengthen false beliefs ([Nyhan & Reifler, 2010](#)), but later, larger studies found this rare ([Wood & Porter, 2019](#)). The bigger problem is the continued influence effect: people keep relying on misinformation even after accepting a correction ([Ecker et al., 2022](#)). The Debunking Handbook recommends leading with the fact, warning before mentioning the myth, explaining the fallacy behind it, and ending by restating the fact ([Lewandowsky et al., 2020](#)).

10. Types of Bias and How to Counter Them

Bias enters research in three ways: from the researcher's mind, from the sources being studied, and from the tools used to find them. Oxford's Centre for Evidence-Based Medicine maintains the Catalogue of Bias, which describes dozens of named biases in health research alone, with definitions and preventive steps (Catalogue [of Bias Collaboration, n.d.](#)). The table below focuses on those most relevant to online research.

10.1 Biases in the Researcher

Confirmation bias, the tendency to seek and interpret evidence to favour existing beliefs, is a ubiquitous phenomenon that appears in many guises ([Nickerson, 1998](#)). It is joined by motivated reasoning, overconfidence (as in the deepfake studies), authority bias, and automation bias, the tendency to trust a machine's output too readily, which [Groh et al. \(2022\)](#) observed when people followed a detection model's mistaken advice.

10.2 Biases in Sources and Media

Sources can be biased through what they choose to cover (selection or story-choice bias), how they frame it (framing and spin), what they leave out (omission), how they sensationalise it, and who pays for it (sponsorship or funding bias). Media rating services such as AllSides, which uses multipartisan editorial panels and blind surveys ([AllSides, n.d.](#)), and Ad Fontes Media, which scores content for reliability and political bias ([Ad Fontes Media, n.d.](#)), can offer a starting point. Their ratings are themselves judgments, so researchers should treat them as one input rather than a verdict.

10.3 Biases in Search Tools

Lesson 31 is particularly strong on this point. Google News personalises results based on search history, region, and language, and a Google News merchandising team selects featured publications and temporary topics for major events ([AOFIRS, n.d.](#)). Search results in the United States differ from those in Canada or Ukraine for the same query, and censorship and conflict zones further filter what is available. Researchers who never change region, language, or account settings are, in effect, researching inside a bubble.

Table 12. Types of bias, how to identify them, and how to counter them

Bias	What it is	How to spot it	How to counter it
Confirmation bias	Favoring evidence that fits what you already believe.	You stop searching once you find agreement.	Search deliberately for the strongest counter-evidence; write down what would change your mind.
Motivated reasoning	Reasoning toward a desired conclusion.	Stricter standards for unwelcome evidence.	Apply identical tests to both sides; ask a colleague to review.

Bias	What it is	How to spot it	How to counter it
Illusory truth	Believing claims because they are familiar.	"I've seen this everywhere."	Count independent origins, not copies.
Overconfidence	Overrating one's own judgment.	Snap verdicts on images and videos.	Rate confidence explicitly; rely on provenance over impression.
Automation bias	Over-trusting tools, detectors or AI answers.	Accepting a tool's verdict without checking.	Treat tool output as a lead; verify independently.
Authority bias	Accepting claims because of titles or logos.	"Experts say" with no names; impressive credentials in an unrelated field.	Check the expert's relevant record and conflicts of interest.
Selection or story-choice bias	What gets covered and what does not.	Some outlets ignore stories others lead with.	Compare coverage across outlets of different leanings and countries.
Framing and spin	Presenting facts to steer interpretation.	Loaded words, selective numbers, emotive images.	Restate the facts in neutral language; find the underlying data.
Sponsorship or funding bias	Results shaped by who pays.	Funder with a stake in the outcome; buried disclosures.	Check funding statements; seek independent replication.
False balance	Treating fringe and consensus views as equal.	Two-sided debates on settled questions.	Check what the weight of evidence and expert bodies say.
Personalization (filter bubble)	Results tailored to your history and profile.	Results match your past interests too neatly.	Use private or signed-out sessions; adjust activity controls as Lesson 31 describes.
Geographic and language bias	Results drawn mainly from your own region and language.	All sources from one country.	Change region parameters and search in the relevant local language.
Recency and ranking bias	Assuming top or newest results are best.	Never looking past page one.	Use date-range operators; dig into archives for the original.

Note. Synthesized from sources (Nickerson, 1998; Kahneman, 2011; Catalogue of Bias Collaboration, n.d.; Lundh et al., 2017; AOFIRS, n.d.).

10.4 A Researcher's Debiasing Routine

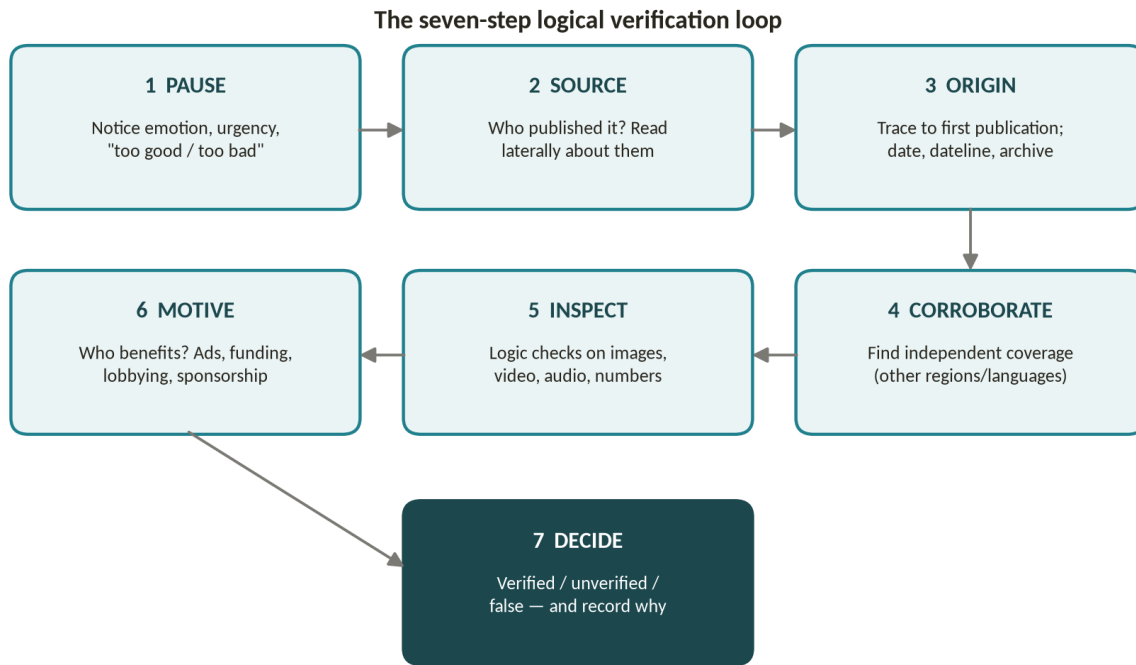
1. **Write the question before searching.** A clear, neutral question resists drift toward the answer you hoped for.
2. **Plan for disagreement.** List the perspectives that should be represented, including those you disagree with, and look for the best version of each.
3. **Triangulate source types.** Combine primary documents, peer-reviewed research, official statistics, and quality journalism rather than relying on one type.
4. **Vary the lens.** Run key searches in another region and language, signed out, and with date ranges.

5. **Separate evidence from inference**—Mark statements that are directly supported and those that are your interpretation.
6. **Log your checks.** Record what you verified, how and when. This makes your work auditable and helps you catch gaps.
7. **Invite a second reader.** A colleague who did not do the research is far better at spotting bias than the author.

11. The Researcher's Practical Toolkit

The following steps bring together all the advice from this report into one routine. For most claims, this process should take just two to five minutes, and only takes longer when the situation is especially important.

Figure 8. The seven-step logical verification loop



Note. This framework combines SIFT ([Caulfield, 2019](#)), lateral reading ([Wineburg & McGrew, 2019](#)), and the timestamp and archive methods in [AOFIRS \(n.d.\)](#).

Table 13. Quick-reference verification checklist

Step	Questions to ask
1. Pause	Am I reacting emotionally? Is this urgent, outrageous or too perfect? Do I actually know this source?
2. Source	Who is behind this? What do independent sources say about them? Are there bylines, bios, contact details and an "About" page that checks out elsewhere?
3. Origin	Where did this first appear? What are the publication, update and event dates? Does an archived version confirm it?
4. Corroborate	Do credible outlets, ideally in more than one country or language, report the same thing independently?
5. Inspect	Do faces, voices, light, shadows, text and backgrounds hold up? Does the setting fit the claimed time and place?
6. Motive	Is anything being sold, promoted or lobbied for? Is there a sponsorship label, affiliate link or funder?

Step	Questions to ask
7. Decide and record	Is it verified, partly verified, unverified or false? What evidence supports that decision? Where is it logged?

12. Limitations of This Report

This report brings together published research and real-world examples, but it did not include new experiments. There are some limitations to keep in mind. Much of the evidence comes from English-speaking, Western countries. Some numbers are from commercial organizations with a product to sell, and we have flagged these. As technology changes, detection studies can become outdated. The search operator methods in Lesson 31 may also have changed, and we have noted any updates we could confirm. In a few cases, like NewsGuard's autumn 2025 count reported by The Ukrainian Week, we could not check the original source, and we have pointed this out.

13. Conclusion

The evidence leads to a clear and hopeful conclusion. The problem is real: false stories spread faster than true ones, synthetic media can trick almost anyone, hidden ads are often missed, and organized groups can create fake agreement or doubt. People are more worried, and trust is dropping.

Yet the solutions are within reach of anyone with a browser and a careful approach. Skills like reading laterally, checking where content comes from, looking at dates and archives, searching in different regions and languages, following the money, and confirming information through independent sources are all free, quick, and easy to learn. Studies show that people improve after just one course, game, or video. For internet research specialists, the goal is to make these habits part of everyday work, be honest about what we do not know, and question our own biases just as much as we question our sources

Closing thought

Seeing is no longer believing. Verifying is.

Appendix A: Query Templates for Verification

The templates below adapt the operator techniques from Lesson 31 to common verification tasks (AOFIRS, n.d.). Replace the bracketed placeholders with your own terms.

Table 14. Query templates for common verification tasks

Task	Template
Has this outlet really published this?	site:[outlet domain] "[exact phrase from the claim]"
When was it first reported?	"[exact phrase]" before:[YYYY-MM-DD]
Find the original version of a story	site:[domain] intitle:"[headline words]" after:[date] before:[date]
Is this text a press release?	"[distinctive sentence]" prnewswire OR businesswire
Who funds this organization?	"[organization name]" funding OR funded OR donors
Is this a sponsored post?	site:[domain] "[article title]" sponsored OR "partner content" OR "paid post"
Find blog discussion only	[topic] inurl:blog inurl:blogs
Exclude a dominant source	[topic] -site:[domain]
Check coverage from another country	Search the translated phrase in the local language; change gl= and hl= parameters in news URLs.

Note. Use quotation marks for exact phrases. Some operators behave differently across search engines and over time.

References

All sources are listed in APA 7th edition style. In-text citations in this report link to the entries below. Web pages without a fixed publication date are shown with the date retrieved.

- Ad Fontes Media. (n.d.). *Methodology*. Retrieved October 7, 2026, from <https://adfontesmedia.com/methodology/>
- Adams, C. (2025, October 31). One trillion web pages archived: Internet Archive celebrates a civilization-scale milestone. *Internet Archive Blogs*. <https://blog.archive.org/2025/10/31/one-trillion-web-pages-archived-internet-archive-celebrates-a-civilization-scale-milestone/>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- AllSides. (n.d.). *Media bias rating methods*. Retrieved October 7, 2026, from <https://www.allsides.com/about/media-bias-rating-methods>
- Amazeen, M. A., & Wojdyski, B. W. (2017). *The effects of native advertising disclosure format on audience perceptions of legacy and online news publishers* [Paper presentation]. Association for Education in Journalism and Mass Communication Annual Conference, Chicago, IL, United States. OpenBU repository. <https://hdl.handle.net/2144/27150>
- An, S., Kerr, G., & Jin, H. S. (2019). Recognizing native ads as advertising: Attitudinal and behavioral consequences. *Journal of Consumer Affairs*, 53(4), 1421–1442. <https://doi.org/10.1111/joca.12235>
- Association of Internet Research Specialists. (n.d.). *Lesson 31: Search in news, articles & blogs* [Unpublished course material, Chapter 1-17]. AOFIRS.
- Bond, S. (2023, May 22). Fake viral images of an explosion at the Pentagon were probably created by AI. *NPR*. <https://www.npr.org/2023/05/22/1177590231/fake-viral-images-of-an-explosion-at-the-pentagon-were-probably-created-by-ai>
- Breakstone, J., Smith, M., Connors, P., Ortega, T., Kerr, D., & Wineburg, S. (2021). Lateral reading: College students learn to critically evaluate internet sources in an online course. *Harvard Kennedy School (HKS) Misinformation Review*, 2(1). <https://doi.org/10.37016/mr-2020-56>
- Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., & Saavedra, A. (2019). *Students' civic online reasoning: A national portrait*. Stanford History Education Group & Gibson Consulting. <https://purl.stanford.edu/gf151tb4868>
- Brewster, J., Effron, E., Fishman, Z., Howard, S., Maitland, E., Sadeghi, M., & Warren, J. (2023, December 27). The year AI supercharged misinformation: NewsGuard's 2023 in review. *Misinformation Monitor*. NewsGuard. <https://www.newsguardtech.com/misinformation-monitor/december-2023/>
- Brown & Williamson Tobacco Corporation. (1969). *Smoking and health proposal* [Internal memorandum]. Truth Tobacco Industry Documents, University of California, San Francisco. <https://www.industrydocuments.ucsf.edu/docs/psdw0147>
- Catalogue of Bias Collaboration. (n.d.). *Catalogue of bias*. Centre for Evidence-Based Medicine, University of Oxford. Retrieved October 7, 2026, from <https://catalogofbias.org/>

- Caulfield, M. (2019, June 19). SIFT (the four moves). *Hapgood*. <https://hapgood.us/2019/06/19/sift-the-four-moves/>
- Coalition for Content Provenance and Authenticity. (n.d.). *C2PA and Content Credentials explainer* (Version 2.3). Retrieved October 7, 2026, from <https://spec.c2pa.org/specifications/specifications/2.3/explainer/Explainer.html>
- Diel, A., Lalgı, T., Schröter, I. C., MacDorman, K. F., Teufel, M., & Bäuerle, A. (2024). Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers. *Computers in Human Behavior Reports*, 16, Article 100538. <https://doi.org/10.1016/j.chbr.2024.100538>
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1, 13–29. <https://doi.org/10.1038/s44159-021-00006-y>
- Edelman, R. (2026, July 22). The Reuters Institute Digital News Report 2026. *Edelman*. <https://www.edelman.com/newsroom/richard-edelmans-6am-blog/reuters-institute-digital-news-report-2026>
- European Commission. (2024, February 14). *Investigation of the Commission and consumer authorities finds that online influencers rarely disclose commercial content* [Press release IP/24/708]. https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_24_708/IP_24_708_EN.pdf
- Farid, H. (n.d.). How to detect faked photos. *American Scientist*. Retrieved October 7, 2026, from <https://www.americanscientist.org/article/how-to-detect-faked-photos>
- Federal Trade Commission. (2015, December 22). *FTC issues enforcement policy statement addressing “native” advertising and deceptively formatted advertisements* [Press release]. <https://www.ftc.gov/news-events/news/press-releases/2015/12/ftc-issues-enforcement-policy-statement-addressing-native-advertising-deceptively-formatted>
- Federal Trade Commission. (2019). *Disclosures 101 for social media influencers*. <https://www.ftc.gov/business-guidance/resources/disclosures-101-social-media-influencers>
- Federal Trade Commission. (2024, August 14). *Federal Trade Commission announces final rule banning fake reviews and testimonials* [Press release]. <https://www.ftc.gov/news-events/news/press-releases/2024/08/federal-trade-commission-announces-final-rule-banning-fake-reviews-testimonials>
- Google. (n.d.). *Refine Google searches*. Google Search Help. Retrieved October 7, 2026, from <https://support.google.com/websearch/answer/2466433?hl=en>
- Granito, J. (2021, June 1). Understanding the sources behind Google News. *Google Search Central Blog*. <https://developers.google.com/search/blog/2021/06/google-news-sources>
- Groh, M., Epstein, Z., Firestone, C., & Picard, R. (2022). Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences*, 119(1), Article e2110013119. <https://doi.org/10.1073/pnas.2110013119>
- Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), Article eaau4586. <https://doi.org/10.1126/sciadv.aau4586>

- Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16(1), 107–112. [https://doi.org/10.1016/S0022-5371\(77\)80012-1](https://doi.org/10.1016/S0022-5371(77)80012-1)
- Holroyd, M., & Olorunselu, F. (2022, March 16). Deepfake Zelenskyy surrender video is the “first intentionally used” in Ukraine war. *Euronews*. <https://www.euronews.com/2022/03/16/deepfake-zelenskyy-surrender-video-is-the-first-intentionally-used-in-ukraine-war>
- iProov. (2025, February 12). *iProov study reveals deepfake blindspot: Only 0.1% of people can accurately detect AI-generated deepfakes* [Press release]. <https://www.iproov.com/press/study-reveals-deepfake-blindspot-detect-ai-generated-content>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kearns, C. E., Schmidt, L. A., & Glantz, S. A. (2016). Sugar industry and coronary heart disease research: A historical analysis of internal industry documents. *JAMA Internal Medicine*, 176(11), 1680–1685. <https://doi.org/10.1001/jamainternmed.2016.5394>
- Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political astroturfing on Twitter: How to coordinate a disinformation campaign. *Political Communication*, 37(2), 256–280. <https://doi.org/10.1080/10584609.2019.1661888>
- Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *iScience*, 24(11), Article 103364. <https://doi.org/10.1016/j.isci.2021.103364>
- 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>
- Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E., Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., ... Zaragoza, M. S. (2020). *The debunking handbook 2020*. Databrary. <https://doi.org/10.17910/b7.1182>
- Lundh, A., Lexchin, J., Mintzes, B., Schroll, J. B., & Bero, L. (2017). Industry sponsorship and research outcome. *Cochrane Database of Systematic Reviews*, (2), Article MR000033. <https://doi.org/10.1002/14651858.MR000033.pub3>
- Magramo, K. (2024, May 16). British engineering giant Arup revealed as \$25 million deepfake scam victim. *CNN Business*. <https://www.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk>
- McGrew, S., Ortega, T., Breakstone, J., & Wineburg, S. (2017). The challenge that’s bigger than fake news: Civic reasoning in a social media environment. *American Educator*, 41(3), 4–9, 39. <https://files.eric.ed.gov/fulltext/EJ1156387.pdf>
- MIT Media Lab. (n.d.). *Detect DeepFakes: How to counteract misinformation created by AI*. Retrieved October 7, 2026, from <https://www.media.mit.edu/projects/detect-fakes/overview/>
- NewsGuard. (2026, June 23). *Tracking AI-enabled misinformation: 3,749 AI content farm sites (and counting), plus the top false claims generated by artificial intelligence tools*. <https://www.newsguardtech.com/special-reports/ai-tracking-center/>

- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, 119(8), Article e2120481119. <https://doi.org/10.1073/pnas.2120481119>
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330. <https://doi.org/10.1007/s11109-010-9112-2>
- Oreskes, N., & Conway, E. M. (2010). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury Press.
- Paris, B., & Donovan, J. (2019). *Deepfakes and cheap fakes: The manipulation of audio and visual evidence*. Data & Society. <https://datasociety.net/research-library/deepfakes-and-cheap-fakes-the-manipulation-of-audio-and-visual-evidence/>
- PBS NewsHour. (2017, October 31). In hearing on Russia, Facebook, Twitter, Google defend their security measures. *PBS NewsHour*. <https://www.pbs.org/newshour/politics/in-hearing-on-russia-facebook-twitter-google-defend-their-security-measures>
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. <https://doi.org/10.1016/j.cognition.2018.06.011>
- Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*, 13, Article 2333. <https://doi.org/10.1038/s41467-022-30073-5>
- Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147(12), 1865–1880. <https://doi.org/10.1037/xge0000465>
- 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>
- Porter, J. (2024, February 2). Google Search’s cache links are officially being retired. *The Verge*. <https://www.theverge.com/2024/2/2/24058985/google-search-cache-feature-discontinued>
- Reuters Institute for the Study of Journalism. (2023). *Overview and key findings of the 2023 Digital News Report*. University of Oxford. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023/dnr-executive-summary>
- Reuters Institute for the Study of Journalism. (2024). *Overview and key findings of the 2024 Digital News Report*. University of Oxford. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024/dnr-executive-summary>
- Reuters Institute for the Study of Journalism. (2025). *Overview and key findings of the 2025 Digital News Report*. University of Oxford. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/dnr-executive-summary>

- Reuters Institute for the Study of Journalism. (2026, June 16). *Overview and key findings of the 2026 Digital News Report*. University of Oxford. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2026/dnr-executive-summary>
- Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5, Article 65. <https://doi.org/10.1057/s41599-019-0279-9>
- 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), Article eabo6254. <https://doi.org/10.1126/sciadv.abo6254>
- Sadeghi, M., & Arvanitis, L. (2023, May 1). *Rise of the newsbots: AI-generated news websites proliferating online*. NewsGuard. <https://www.newsguardtech.com/special-reports/newsbots-ai-generated-news-websites-proliferating>
- Shepardson, D. (2024, September 26). Consultant fined \$6 million for using AI to fake Biden's voice in robocalls. *Reuters*. <https://www.reuters.com/world/us/fcc-finalizes-6-million-fine-over-ai-generated-biden-robocalls-2024-09-26/>
- Silverman, C. (Ed.). (2014). *Verification handbook: A definitive guide to verifying digital content for emergency coverage*. European Journalism Centre. <http://verificationhandbook.com/>
- Sleek, S. (2022, April 29). Vaccinating against bunk: Curbing viral misinformation. *APS Observer*. Association for Psychological Science. <https://www.psychologicalscience.org/observer/vaccinating-against-bunk>
- Southern, M. G. (2021, June 1). Google on how it ranks news sources. *Search Engine Journal*. <https://www.searchenginejournal.com/google-on-how-it-ranks-news-sources/409118/>
- Stanford History Education Group. (2016). *Evaluating information: The cornerstone of civic online reasoning* [Executive summary]. Stanford University. <https://stacks.stanford.edu/file/druid:fv751yt5934/SHEG%20Evaluating%20Information%20Online.pdf>
- Stocking, G., Wang, L., Lipka, M., Matsa, K. E., Widjaya, R., Tomasik, E., & Liedke, J. (2024, November 18). *America's news influencers: The creators and consumers in the world of news and information on social media*. Pew Research Center. <https://www.pewresearch.org/journalism/2024/11/18/americas-news-influencers/>
- The Ukrainian Week. (2026, August 10). From fake news to fake sources: Russia's push to influence AI. *The Ukrainian Week*. <https://ukrainianweek.com/russia-fake-news-ai/>
- van der Linden, S. (2023). *Foolproof: Why misinformation infects our minds and how to build immunity*. W. W. Norton.
- 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>
- Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, 37(3), 350–375. <https://doi.org/10.1080/10584609.2019.1668894>
- 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.

<https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>

WebCite. (n.d.). *WebCite frequently asked questions*. Retrieved October 7, 2026, from

<https://www.webcitation.org/faq>

Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record*, 121(11), 1–40.

<https://doi.org/10.1177/016146811912101102>

Wineburg, S., Breakstone, J., McGrew, S., Smith, M. D., & Ortega, T. (2022). Lateral reading on the open Internet: A district-wide field study in high school government classes. *Journal of Educational Psychology*, 114(5), 893–909. <https://doi.org/10.1037/edu0000740>

Wojdyski, B. W., & Evans, N. J. (2016). Going native: Effects of disclosure position and language on the recognition and evaluation of online native advertising. *Journal of Advertising*, 45(2), 157–168.

<https://doi.org/10.1080/00913367.2015.1115380>

Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence.

Political Behavior, 41(1), 135–163. <https://doi.org/10.1007/s11109-018-9443-y>

Index

Page numbers refer to the main body of the report (Executive Summary through Appendix A).

A

Accuracy prompts, 24–25
 Ad Fontes Media, 30
 Advertorials, 5, 20, 23
 Affiliate links and reviews, 5, 20, 23, 33
 AI chatbots, 8, 29
 AI content farms, 8–9, 23, 29
 AllSides, 30
 Archives, web, 3, 12–13, 17, 27–29, 31, 33, 35–36
 Arup video-call fraud, 2, 18, 26
 Associated Press, 12–13
 Astroturfing, 6
 Authority bias, 30–31
 Automation bias, 30–31

B

Backfire effect, 29
 Bad News game, 24–25, 27
 Biden robocall, 26
 Blogs, searching, 1, 3, 12, 20, 27, 37

C

C2PA (Content Credentials), 17
 Cache, Google, 13, 29
 Cheap fakes, 5, 17
 Confirmation bias, 30
 Continued influence effect, 29
 Corroboration, 17–18, 26–28, 33
 CRAAP checklist, 28

D

Datelines, 3, 11–12
 Debunking Handbook, 29
 Deepfakes, 1–5, 11–12, 14–18, 23–24, 26, 30
 Disinformation, 5
 DMOZ directory, 13, 27

E

Emotion and virality, 7, 11, 17, 33
 European Commission, 1, 3, 19–21

F

Fact-checking, 10, 17, 23–25
 False balance, 31
 Federal Trade Commission (FTC), 3, 5, 20–21, 23
 Filter bubble, 13, 28, 30–31

Framing and spin, 30–31
 Front groups, 6, 22–23, 26
 Funding bias, 22, 26, 30–31

G

Google News, 13, 30

I

Illusory truth effect, 7, 29, 31
 Influencers, 1, 5, 7, 20–21, 23
 Inoculation (prebunking), 24–25, 27
 intitle: operator, 12, 26, 37
 inurl: operator, 11–12, 26–27, 29, 37

L

Language and translation, 13, 27–28, 31, 37
 Lateral reading, 10, 20, 23–26, 28, 33, 36
 Liar's dividend, 18
 Lobbying, 3, 11, 22–23

M

Malinformation, 5
 Manufactured doubt, 6, 22
 MinimumWage.com, 11, 22, 26
 Misinformation, 5, 29
 Motivated reasoning, 7, 30

N

Native advertising, 1, 5, 19, 23–24
 NewsGuard, 4, 8–9, 23, 29, 35
 Newswires, 3, 12–13, 20, 23, 37

O

Out-of-band confirmation, 18, 26, 28
 Overconfidence, 1, 14, 24, 30–31

P

Pentagon explosion image, 26
 Personalization of search, 29–31
 Pew Research Center, 3, 23
 Press releases, 5, 7, 13, 20, 23, 37
 Provenance, 2, 17–18, 31

R

Region and location settings, 3, 13, 23, 27–28, 30–31, 36
 Reuters Institute Digital News Report, 1, 3, 7–8
 Reverse image search, 17, 28

S

Search operators, 3, 11–13, 27–29, 31, 35, 37
 Selection bias, 30
 SIFT method, 10, 28, 33

site: operator, 11–12, 37

Sponsored content, 1, 5, 19, 37

Sugar industry, 22, 26

T

Timestamps, 3, 12, 28, 33

Tobacco industry, 22

Truth sandwich, 29

U

URL structure, 3, 11–12, 27–29

V

Verification debt, 29

Verification scale, 28, 34

W

Wayback Machine, 12–13, 17, 27

WebCite, 13, 29

Z

Zelensky deepfake, 26