IPIE

International Panel on the Information Environment

Responding to Generative AI Misinformation

Results from a Meta-Analysis of Scientific Evidence

Summary for Policymakers 2026.2

DOI Number: 10.61452/QXAF2136

A. Herasimenka, S. Valenzuela, S. Boulianne, F. Esser, L. M. Given, S. Lewandowsky, E. M. Navarro-Lopez, P. N. Howard

IPIE
International Panel on the Information Environment

Responding to Generative AI Misinformation

Results from a Meta-Analysis of Scientific Evidence

Summary for Policymakers 2026.2

How to cite:
International Panel on the Information Environment [A. Herasimenka, S. Valenzuela, S. Boulianne, F. Esser, L. M. Given, S. Lewandowsky, E. M. Navarro-López, P. N. Howard (eds.)], "Responding to Generative AI Misinformation: Results from a Meta-Analysis of Scientific Evidence," Zurich, Switzerland: IPIE, 2026. Summary for Policymakers, SFP2026.2, doi: 10.61452/QXAF2136.

SYNOPSIS

Generative artificial intelligence (GenAI) can now rapidly produce large volumes of misleading text, images, audio, and video with readily accessible tools. This matters for policymakers because false or misleading content can be tailored to specific audiences. Such content can spread quickly and mimic authentic communication, making it difficult for people to judge what is true.

This Summary for Policymakers distills the main findings of the IPIE Synthesis Report (SR2026.2) on the effects of misinformation produced with GenAI and the measures most likely to reduce its influence.

The assessment draws on a large-scale meta-analysis of experimental scientific evidence. It is based on 60 randomized controlled trial effect estimates from 24 peer-reviewed publications involving 33,801 participants, published between 2018 and 2025.

The report reaches four main conclusions:

1. Text-based GenAI misinformation currently poses greater persuasive risks than visual misinformation.
2. The current evidence base excludes most of the world. Research is concentrated in English-language and high-income countries, leaving significant gaps in our knowledge.
3. The most consistently effective intervention is to provide users with corrective information, preventively, so they can evaluate accuracy and credibility themselves.
4. Labeling is effective when it is consistent. Content labeling generally decreases perceived credibility, but its impact varies greatly depending on how the firms create and apply labels.

The policy implications are clear. Policymakers should prioritize addressing misinformation from text-based GenAI and support preventive and corrective information as a fundamental strategy. Labeling should be treated as a measure that requires careful design and testing. We must expand independent research well beyond English-language and high-income contexts.

Researcher access to platform and model data is crucial for enhancing public understanding of GenAI misinformation and for testing which safeguards are effective in practice.

INTRODUCTION

Generative artificial intelligence (GenAI) systems can produce text, images, audio, and videos that are sometimes hard for people to tell apart from content created by humans. AI-generated content has become widespread, appearing in elections worldwide—often from unknown or malicious sources—and increasingly influencing public perception in areas such as public health, conflict, and markets, with significant real-world effects [1]. These systems include large language models that generate text, image and audio synthesis tools, and video creation systems. They can also produce misleading or harmful content, such as fake news, rumors, propaganda, or deepfakes. In this report, these phenomena are collectively referred to as misinformation.

This Summary for Policymakers highlights the main findings of a meta-analysis of experimental evidence on the effects on individuals of exposure to misinformation created by GenAI [2]. It also assesses two countermeasures supported by enough experimental evidence for policy discussion: corrective information and content labeling. The report emphasizes randomized controlled trials because these studies offer more reliable comparisons across experiments and allow for more systematic estimation of average effects.

The evidence shows a rapidly evolving information environment. The report finds that public responses to text and visual misinformation are now diverging. It also finds that the evidence base remains limited in its geographic scope and methods. At the same time, it identifies two practical responses that can reduce the perceived credibility of GenAI misinformation, although one is more consistent than the other.

To build this evidence base, the review searched two major bibliographic databases, Scopus and Web of Science, incorporated expert recommendations, merged the results, removed duplicates, and screened a final set of 6,952 publications. After eligibility screening and full-text coding, 87 publications met the criteria for detailed review. Additional bias checks, screening, and protocol reviews following Cochrane recommendations resulted in a subset of 24 particularly relevant studies for inclusion in the meta-analysis. The main Synthesis Report employs three-level random-effects meta-analysis and evaluates the risk of bias using the Revised Tool for Risk of Bias in Randomized Trials. Please see the full Synthesis Report [2] for complete methodology details.

RESULT 1. MISLEADING GENAI TEXT IS THE BIGGEST THREAT

KEY FINDING

Text-based generative AI misinformation currently poses greater persuasive risks than visual misinformation.

The report reveals a clear difference between the perception of trustworthiness between text and visual GenAI content. More recent studies show that false GenAI-generated text is often perceived as more accurate, credible, or believable than true information. At the same time, studies on false GenAI-generated visual information, such as deepfakes, are often perceived as less credible than true information. In visual studies conducted after 2021, the pooled estimate indicates a moderate drop in perceived credibility.

This does not mean that visual misinformation is safe or unimportant. Deepfakes still pose risks, and the report notes that public responses may change again as visual systems improve. However, the current evidence indicates that text-based misinformation deserves a more urgent policy focus. This is significant because text-focused systems can be inexpensive, scalable, easy to customize, and increasingly integrated into search, messaging, and conversational interfaces. The report also indicates that the outputs of conversational AI may be significantly more persuasive than those of static AI-generated messages, suggesting that current estimates may be conservative.

The overall pattern of findings is strong enough to support a clear policy message: the most immediate and persuasive risk currently stems from false or misleading text, not from visual deepfakes alone.

RESULT 2. GLOBAL EVIDENCE GAPS

The report clearly emphasizes that most of the examined publications focus on English-language and high-income countries, such as Austria, Germany, the Netherlands, South Korea, the United Kingdom, and the USA. The current evidence is limited to full-text articles in English, and five screened publications in other languages did not meet the eligibility criteria. Consequently, current understanding of GenAI misinformation is concentrated in a small number of countries, languages, and media environments.

This is a significant limitation. It indicates that policymakers, regulators, researchers, and companies still have limited knowledge about how GenAI misinformation operates in many other regions. The report explicitly states that

KEY FINDING

The current evidence base leaves most of the world—countries and languages—out of scope.

this creates a substantial knowledge gap beyond the few countries and languages that are publishing on this topic. This gap is important for multilingual societies, cross-border regulation, and efforts to develop effective safeguards across different languages, cultures, and political systems.

Additionally, the evidence is limited by the rapid pace of technological advancement. Many experiments test older systems such as GPT-2 or early tools like FaceSwap. The report warns that scientific research often lags behind current technology. As a result, policies are sometimes based on evidence from weaker systems than those currently in use publicly. Therefore, independent research that examines broader contexts is a practical necessity to ensure that regulation and governance remain aligned with the technology shaping public discourse.

RESULT 3. PREVENTIVE CORRECTIONS CAN HELP

Preventive corrective information includes brief educational warnings, reminders or prompts, and other ways of inoculating users before exposure to potentially misleading content. For example, some interventions provide short explanations of how deepfakes enable realistic video manipulation or remind users that GenAI systems can produce errors. These approaches aim to improve users' ability to assess whether content is trustworthy.

KEY FINDING

Preventive corrective information is the most consistently effective intervention.

In post-2020 studies, the pooled estimate shows that preventive corrective information reduces perceptions of accuracy, credibility, or acceptance of GenAI misinformation, with a small to moderate effect size. These findings are relatively consistent across the reviewed studies: variation between their results is low, and effects are observed across different populations and experimental designs in this subset of evidence.

These findings indicate that such interventions are likely to remain effective in the near term, although most of the available evidence is based on studies of earlier generations of GenAI systems.

The Synthesis Review also highlights an important caveat: the positive impact of preventive corrective information appears when people are asked to evaluate content for accuracy or credibility. The evidence is less consistent when studies focus only on detection. Earlier research on deepfakes found that warnings could sometimes increase distrust of both real and manipulated videos. This aligns with a broader concern in misinformation research and policy: false narratives and propaganda often aim at making an

individual distrust all public information. Hence, interventions should be designed to improve individuals' ability to distinguish between reliable and misleading content.

Preventive corrective information is one of the most widely tested measures that can be deployed online relatively quickly, especially on social media platforms. Examples include notes stating that GenAI can produce errors, explanations of how deepfakes enable realistic video manipulation, and brief educational materials on phenomena like AI hallucinations. Depending on the discipline, such measures are referred to as advice, inoculation, pre-bunking, priming, or warning labeling.

RESULT 4. LABELING WORKS WHEN CONSISTENT

KEY FINDING

Content labeling remains promising, but only when it is clear and consistent.

Content labeling refers to the use of short visual, textual, or multimodal tags attached to information. These labels indicate that the information may have been created with GenAI, has been altered, or be misleading. In the pooled analysis, labels are linked to a small but statistically significant decrease in the perceived credibility of misleading GenAI information. This is an encouraging result, but it is not consistent across all cases. The report finds considerably more variation in labeling evidence than in evidence on preventive corrective information.

Because of limited evidence, there is considerable variation in effects, and some studies find no effects at all. Labels might decrease credibility in some situations, but not in others. This assessment emphasizes likely moderating factors, including label design, wording, presentation, content type, experimental setting, and label source. In other words, how labels are designed and implemented matters at least as much as whether labels are used at all.

The evidence base for labeling is also limited. The studies in this relatively small sample focus almost entirely on participants in the USA. That makes it difficult to know how labels will perform across different languages, cultures, or legal systems. Therefore, the right conclusion is not that labeling fails. Instead, labeling can be helpful, but only when firms consider the design of these tags and test them across various contexts rather than assuming they work everywhere by default.

CONCLUSION

This Summary for Policymakers highlights a shifting pattern of risk. Misinformation created by text-based GenAI poses a greater persuasive threat than visual-based GenAI. However, this does not eliminate the dangers of deepfakes or other synthetic media. It indicates that policymakers should not let the prominence of visual deception divert attention from the growing persuasive power of false AI-generated text.

This report emphasizes the results of previous assessments by IPIE [3], [4] and suggests two practical responses. Preventive corrective information is the most consistent and well-supported intervention in the current evidence base, especially when provided before exposure and when it enhances better evaluative judgment. Content labeling can also be helpful, but its effectiveness varies and depends on context. Labels are most effective when carefully designed and applied clearly and consistently.

The final point concerns the evidence itself. Most of the world remains outside the scope of current research. This limits policymakers, platforms, and developers of general-purpose AI models' ability to understand how these risks propagate across languages, cultures, and media systems. Therefore, independent research, ongoing evidence synthesis, and improved researcher access to platform and model data are essential. Without these, public policy will remain reactive, incomplete, and too narrowly focused in a rapidly evolving information environment.

REFERENCES

[1] International Panel on the Information Environment [I. Trauthig, P. N. Howard, S. Valenzuela (eds.)], "The Role of Generative AI Use in 2024 Elections Worldwide," Zurich, Switzerland: IPIE, 2025. Technical Paper, TP2025.2, doi: 10.61452/HZUE9853.

[2] International Panel on the Information Environment [A. Herasimenka, S. Valenzuela, S. Boulianne, F. Esser, L. M. Given, S. Lewandowsky, E. M. Navarro-López, P. N. Howard (eds.)], "Effects of Misinformation Produced with Generative AI and its Countermeasures: A Meta-Analysis of Experimental Scientific Evidence," Zurich, Switzerland: IPIE, 2026. Synthesis Report, SR2026.2, doi: 10.61452/UGTR3022.

[3] International Panel on the Information Environment, "Countermeasures for Mitigating Digital Misinformation: A Systematic Review," IPIE, Zurich, Switzerland, SR2023.1, July 2023. [Online]. Available: https://www.ipie.info/research/sr2023-1

[4] International Panel on the Information Environment, "Platform Responses to Misinformation: A Meta-Analysis of Data," IPIE, Zurich, Switzerland, SR2023.2, July 2023. [Online]. Available: https://www.ipie.info/research/sr2023-2

ACKNOWLEDGMENTS

Contributors

Drafting authors: Aliaksandr Herasimenka (Consulting Scientist, United Kingdom), Sebastián Valenzuela (IPIE Chief Science Officer and Chair of the Science & Methodology Committee, Chile), Shelley Boulianne (IPIE Science & Methodology Committee Member, Canada), Frank Esser (IPIE Science & Methodology Committee Member, Switzerland), Lisa M. Given (IPIE Science & Methodology Committee Member, Canada/Australia), Stephan Lewandowsky (IPIE Science & Methodology Committee Member, Australia/United Kingdom), Eva M. Navarro-López (IPIE Science & Methodology Committee Member, Spain/UK/Mexico), Philip Howard (IPIE President and CEO, Canada/UK). Research Assistants: Anna George and Xianlingchen Wang. Independent General Reviews: George Georgarakis and Mathias Harrer. Design: Domenico Di Donna. Copyediting: Beverley Sykes. We gratefully acknowledge support from the IPIE Secretariat: Lola Gimferrer, Jessica Gold, Wiktoria Schulz, Donna Seymour, Anna Staender, and Alex Young.

Funders

The International Panel on the Information Environment (IPIE) gratefully acknowledges the support of its funders. For a full list of funding partners please visit www.ipie.info. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the IPIE and do not necessarily reflect the views of the funders.

Declaration of Interests

IPIE reports are developed and reviewed by a global network of research affiliates and consulting scientists who constitute focused Scientific Panels and contributor teams. All contributors and reviewers complete declarations of interests, which are reviewed by the IPIE at the appropriate stages of work.

Preferred Citation

An IPIE *Summary for Policymakers* provides a high-level precis of the state of knowledge and is written for a broad audience. An IPIE *Synthesis Report* makes use of scientific meta-analysis techniques, systematic review, and other tools for evidence aggregation, knowledge generalization, and scientific consensus building, and is written for an expert audience. An IPIE *Technical Paper* addresses particular questions of methodology, or provides a policy analysis on a focused regulatory problem. All reports are available on the IPIE website (www.IPIE.info).

This document should be cited as:

International Panel on the Information Environment [A. Herasimenka, S. Valenzuela, S. Boulianne, F. Esser, L. M. Given, S. Lewandowsky, E. M. Navarro-López, P. N. Howard (eds.)], "Responding to Generative AI Misinformation: Results from a Scientific Meta-Analysis," Zurich, Switzerland: IPIE, 2026. Summary for Policymakers, SFP2026.2, doi: 10.61452/QXAF2136.

Copyright Information

ABOUT THE IPIE

The International Panel on the Information Environment (IPIE) is an independent and global science organization committed to providing the most actionable scientific knowledge about threats to the world's information environment. Based in Switzerland, the mission of the IPIE is to provide policymakers, industry, and civil society with independent scientific assessments on the global information environment by organizing, evaluating, and elevating research, with the broad aim of improving the global information environment. Hundreds of researchers from around the world contribute to the IPIE's reports.

For more information, please contact the International Panel on the Information Environment (IPIE), secretariat@IPIE.info. Seefeldstrasse 123, P.O. Box, 8034 Zurich, Switzerland.

International Panel on
the Information
Environment

Seefeldstrasse 123
P.O. Box 8034 Zurich
Switzerland