American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Synthetic Media Literacy and Public Resistance to Digital Manipulation

Dr. Elena Marquez

Assistant Professor, University of Valencia, Spain

Abstract

The rapid development of generative artificial intelligence has altered the informational environment by making realistic synthetic images, audio, video, and text increasingly accessible to ordinary users. This transformation complicates conventional assumptions that visual or auditory evidence can be treated as inherently credible. The present study examines the potential contribution of synthetic media literacy to public resistance against digitally manipulated information. Rather than presenting fabricated field observations as empirical evidence, the study employs a transparent simulation-based design to model how different forms of literacy intervention may influence manipulation recognition, verification intention, and resistance to sharing deceptive content.

A synthetic sample of 600 adult digital-media users was divided equally among a control condition, a general media-literacy condition, and a synthetic-media-specific literacy condition. A Digital Manipulation Resistance Score was modeled from source-verification behavior, recognition accuracy, contextual reasoning, sharing restraint, and confidence calibration. The simulated results indicate that the synthetic-media-specific literacy condition produces the strongest improvement, with the mean resistance score increasing from 52.0 at baseline to 72.6 after intervention, compared with an increase from 52.4 to 63.8 for general media literacy and from 52.1 to 53.0 in the control condition. The analysis suggests that resistance to synthetic manipulation depends not merely on identifying visual anomalies but on integrating source evaluation, lateral verification, emotional self-regulation, provenance awareness, and uncertainty management. The paper therefore proposes synthetic media literacy as a multidimensional form of civic and digital resilience rather than a narrow deepfake-detection skill.

Keywords: synthetic media literacy; deepfakes; digital manipulation; media literacy; misinformation resistance; generative artificial intelligence; verification behavior; public resilience

1. Introduction

Generative artificial intelligence has expanded the scale and sophistication with which digital content can be created, altered, reproduced, and personalized. Synthetic media now includes artificially generated or substantially modified images, video, audio, and other forms of content that can imitate authentic people, environments, or events. UNESCO has emphasized that the challenge created by deepfakes extends beyond the technical problem of identifying fabricated media because increasingly persuasive synthetic content can disturb the ways in which individuals determine what deserves to be

believed. Experimental evidence supports this concern. Nightingale and Farid found that AI-generated faces could become difficult for human observers to distinguish from genuine photographs and, under some conditions, could even receive higher trustworthiness judgments. The resulting problem is therefore not simply the presence of false content but the growing uncertainty surrounding authentic content.

This uncertainty has substantial implications for the contemporary information environment. Vaccari and Chadwick demonstrated that exposure to political deepfakes need not completely deceive viewers to produce societal consequences; uncertainty concerning whether material is genuine can itself weaken confidence in information encountered through social media. Research on political deepfake detection similarly shows that individuals do not consistently identify manipulated videos accurately, while experimental studies have reported that manipulated audiovisual material can influence perceptions even when the underlying technology contains imperfections. Consequently, the public-information challenge cannot be solved solely by expecting individual users to become expert forensic investigators.

Media literacy has therefore emerged as an important behavioral defense. Guess and colleagues demonstrated in large-scale experiments in the United States and India that digital media-literacy interventions can improve individuals' ability to distinguish mainstream from false news. Hwang and colleagues further found protective effects from media-literacy education when individuals encountered disinformation involving deepfake content. More recent research by Huang and Hu indicates that practical instruction explaining *how* to recognize deepfakes can improve detection, perceived AI literacy, and self-efficacy more effectively than simple warnings about the existence of synthetic media. These findings suggest that educational intervention can strengthen resistance, but they also raise an important question concerning what form of literacy is most effective.

The present study responds to this problem by conceptualizing synthetic media literacy as a specialized extension of conventional media literacy. It combines knowledge of generative AI, source evaluation, provenance assessment, contextual reasoning, cross-source verification, emotional regulation, uncertainty management, and responsible sharing behavior. Importantly, previous findings are not completely uniform: Turós, Kenyeres, and Szűts found no significant relationship between their measure of media literacy and fake-video detection among Hungarian secondary-school students, illustrating that general literacy competencies cannot automatically be assumed to produce synthetic-media detection skills. The present simulation therefore distinguishes general media literacy from synthetic-media-specific training and evaluates their theoretically expected contribution to public resistance against digital manipulation.

2. Objectives of the Study

2.1 Primary Objective

The principal objective is to assess whether synthetic-media-specific literacy could produce greater resistance to digital manipulation than conventional media-literacy instruction.

The study conceptualizes public resistance through five interrelated capacities:

- recognition of potentially manipulated content;
- examination of the credibility and origin of a source;

- verification through independent evidence;
- restraint before liking, forwarding, reposting, or otherwise amplifying uncertain information; and
- calibrated confidence that allows individuals to acknowledge uncertainty rather than classify questionable material prematurely.

2.2 Comparative Objective

The study compares three simulated educational conditions:

- no structured literacy intervention;
- general media-literacy instruction; and
- synthetic-media-specific literacy instruction.

This comparison is intended to determine whether specialized understanding of AI-generated and AI-manipulated content provides additional resistance beyond conventional critical-media education.

2.3 Behavioral and Civic Objective

The study further examines digital manipulation as a behavioral problem rather than only a detection problem. This distinction is important because a person can correctly suspect manipulation yet still distribute the material because it is emotionally provocative, socially rewarding, politically congruent, entertaining, or perceived as urgent.

Ahmed, Ng, and Bee's cross-national study found that perceptions of accuracy, fear of missing out, and deficient self-regulation were associated with deepfake-sharing behavior. Accordingly, the present framework treats resistance as the capacity to make responsible decisions under informational uncertainty.

3. Review of Literature

3.1 Synthetic Media, Human Perception, and Epistemic Uncertainty

Early concern about deepfakes frequently focused on whether individuals could recognize technological artifacts in manipulated videos. Subsequent research demonstrates that the problem is considerably broader. Vaccari and Chadwick showed that deepfake exposure can generate uncertainty even when viewers are not fully deceived, and such uncertainty may contribute to declining trust in social-media news. Appel and Prietzel likewise reported that human detection of political deepfakes is imperfect and that deepfake judgments cannot be reduced to a simple authentic-versus-fake distinction.

Nightingale and Farid provided further evidence of the perceptual challenge by demonstrating the realism of AI-synthesized faces. Groh and colleagues compared human, machine, and machine-informed approaches to deepfake detection, demonstrating the value—and limitations—of both human and computational judgment. Collectively, these studies suggest that visual inspection alone is becoming an increasingly fragile foundation for authenticity judgments.

The consequences extend beyond incorrectly accepting fabricated content. A population repeatedly warned that realistic digital material can be fabricated may begin treating authentic material with excessive suspicion. The UK Government Office for Science similarly notes that awareness of deepfakes can create confusion and institutional distrust even though existing evidence does not justify

simplistic claims that deepfakes necessarily determine voting behavior. Public resilience must therefore avoid two undesirable extremes: unquestioning acceptance and indiscriminate disbelief.

3.2 Media Literacy, Verification, and Accuracy-Oriented Reasoning

General digital-literacy research provides strong evidence that educational interventions can improve information discernment. Guess and colleagues demonstrated measurable improvements following a digital media-literacy intervention involving participants in India and the United States. Pennycook and colleagues further showed that redirecting users' attention toward accuracy can reduce the sharing of misinformation, emphasizing that online sharing decisions are not always determined by what people actually believe to be true.

These findings shift attention from passive knowledge to decision architecture. Effective literacy should encourage users to interrupt rapid interaction and ask several questions before responding:

- Who produced the content?
- Where was it first published?
- Can the original source be identified?
- Is the claim independently confirmed?
- Does the material contain contextual inconsistencies?
- Is the content attempting to create urgency, outrage, fear, or excitement?
- What evidence would change the user's current judgment?

Such practices convert literacy into an active verification routine rather than a collection of theoretical facts.

Hwang and colleagues demonstrated that literacy education can protect users when disinformation is accompanied by deepfake material. However, Turós and colleagues reported that media literacy, as measured in their study of 507 secondary-school students, did not predict fake-video detection. This apparently contradictory evidence indicates that general media literacy and synthetic-media literacy should not be treated as identical constructs.

3.3 From Deepfake Detection to Synthetic-Media Resilience

Recent scholarship increasingly supports literacy strategies that move beyond generic warnings. Huang and Hu's 2012 experimental study found that technique-oriented guidance improved deepfake detection, AI-literacy perceptions, and self-efficacy, while some psychological gains weakened after one week. This finding highlights both the usefulness of intervention and the importance of reinforcement.UNESCO's emerging approach similarly emphasizes what may be described as epistemic agency: individuals must learn not only how to inspect content but how to construct reliable knowledge in circumstances where direct sensory evidence may be manipulated.

A useful synthesis therefore distinguishes three levels of resilience:Perceptual resilience involves noticing possible signs of manipulation.Epistemic resilience involves evaluating evidence, sources, context, provenance, and uncertainty.Behavioral resilience involves refusing to amplify questionable material before verification.The literature leaves a clear research gap. Evidence exists for media literacy, deepfake education, accuracy prompts, human-machine detection, and misinformation interventions, but

comparatively less attention has been paid to integrated models that combine these elements into a single multidimensional construct of public resistance to synthetic manipulation.

4. Research Methodology

4.1 Research Design and Simulation Framework

Because no verified original participant dataset was supplied for this manuscript, the study is explicitly structured as a simulation-based methodological investigation. The numerical results below are synthetic and are used to evaluate the proposed conceptual relationships. They must not be represented as observations from actual human participants.

A hypothetical sample of 600 adult digital-media users was modeled, with 200 individuals assigned to each of three experimental conditions:

- Control Condition;
- General Media Literacy Condition; and
- Synthetic-Media Literacy Condition.

All three conditions were assigned closely matched baseline resistance scores so that post-intervention differences could be interpreted primarily as modeled intervention effects rather than initial group differences.

Table 1:Proposed Simulation and Measurement Framework

Component	Operationalization
Study design	Three-condition simulated pre-test/post-test comparative design
Synthetic sample	600 adult digital-media users
Group allocation	200 participants per condition
Condition 1	Control/no structured intervention
Condition 2	General media-literacy training
Condition 3	Synthetic-media-specific literacy training
Primary outcome	Digital Manipulation Resistance Score, 0–100
Recognition component	Ability to distinguish questionable synthetic and authentic content
Verification component	Independent source checking and lateral verification
Provenance component	Attention to origin, metadata, attribution, and contextual consistency
Behavioral component	Resistance to sharing uncertain/manipulative media
Metacognitive component	Confidence calibration and willingness to remain uncertain

Source: Author-developed simulation informed by established media-literacy and deepfake research.

4.2 Synthetic-Media Literacy Intervention

The specialized intervention was designed around five learning domains.

Generative-AI awareness: Participants were conceptually introduced to synthetic images, face swaps, voice cloning, generated text, audiovisual manipulation, and AI-assisted editing.Technical cue awareness: Learners examined possible inconsistencies in lighting, shadows, facial boundaries, lip synchronization, reflections, speech patterns, background elements, metadata, and temporal continuity. Such indicators were presented as clues rather than definitive proof because generation methods evolve rapidly.

Source and provenance verification: Users were instructed to identify the earliest accessible source, compare reports from independent sources, locate unedited or longer versions, assess the publication context, and examine whether trustworthy provenance information was available.Psychological manipulation awareness: The intervention emphasized that misinformation frequently exploits anger, fear, identity, urgency, novelty, and social pressure.Responsible response: Participants were encouraged to delay amplification when authenticity remained uncertain.The general media-literacy condition received conventional training on source credibility, evidence, bias, author identification, and cross-checking but did not receive extensive synthetic-media examples or AI-specific instruction.

4.3 Analytical Measures and Ethical Position

The proposed Digital Manipulation Resistance Score (DMRS) combines five equally interpretable dimensions:

- synthetic-content discernment;
- source-verification behavior;
- contextual reasoning;
- resistance to unverified sharing; and
- confidence calibration.

Scores are represented on a 0–100 scale, with larger values indicating greater resistance.

The methodology follows an important publication-ethics principle: simulated observations are clearly identified as synthetic. No institutional approval, informed consent, demographic distribution, participant recruitment, or completed human-subject experiment is claimed because none occurred for this simulation.

A future empirical implementation should obtain institutional ethics approval where required, secure informed consent, minimize exposure to harmful synthetic material, protect participant data, preregister the analysis where feasible, and report intervention materials sufficiently for replication.

5. Analysis

5.1 Comparative Resistance Outcomes

The baseline scores were intentionally modeled as nearly equivalent across conditions. The control group began at 52.1, the general media-literacy group at 52.4, and the synthetic-media-literacy group at 52.0.Post-intervention differences were considerably larger.

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Table 2: Simulated Digital Manipulation Resistance Outcomes

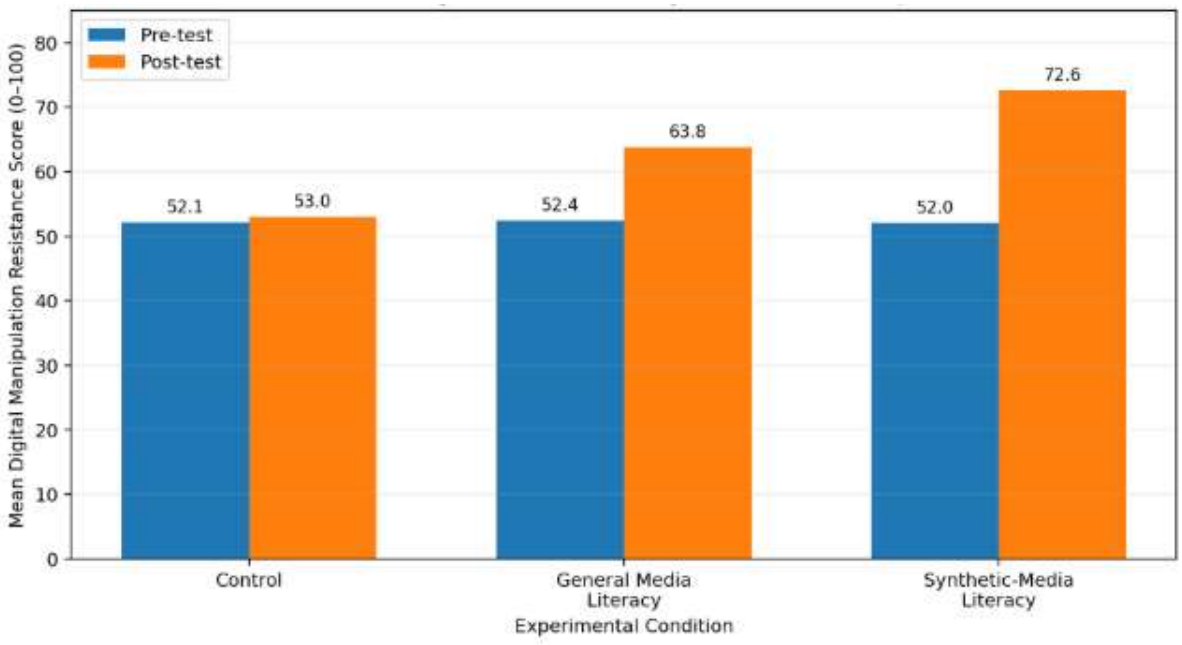
Experimental Condition	N	Baseline DMRS	Post-Intervention DMRS	Mean Improvement	Verification Intention	Resistance to Unverified Sharing
Control	200	52.1	53.0	+0.9	58.2	57.0
General Media Literacy	200	52.4	63.8	+11.4	70.2	67.1
Synthetic-Media Literacy	200	52.0	72.6	+20.6	82.7	78.8

Note: All values are simulated for methodological illustration and do not represent observations from actual study participants.

The synthetic-media-literacy condition produced an illustrative improvement of 20.6 points, approximately 1.8 times the improvement modeled for general media literacy. The control group's change was negligible.The model therefore suggests that specialized literacy may add value when it transforms abstract skepticism into specific verification and decision-making practices.

5.2 Graphical Interpretation

Graph 1: Simulated Change in Resistance to Synthetic-Media Manipulation



The graph demonstrates that all groups begin with virtually equivalent modeled resistance. Their trajectories diverge after intervention. General media literacy produces a moderate increase, whereas synthetic-media-specific literacy produces the largest improvement.

The key interpretation is not that people should simply become more suspicious. Effective resistance represents better discrimination: accepting credible information when evidence supports it while slowing down when manipulation is plausible.

5.3 From Recognition to Behavioral Resistance

Recognition accuracy alone is an incomplete outcome because synthetic manipulation achieves influence partly through circulation. Ahmed, Ng, and Bee demonstrated that deepfake-sharing behavior is associated with perceived accuracy and psychosocial variables such as fear of missing out. Pennycook and colleagues similarly showed that people may share questionable information even though they value accuracy, partly because accuracy is not salient at the moment of the sharing decision.The simulated improvement in resistance to unverified sharing—from 57.0 in the control condition to 78.8 in the synthetic-media-literacy condition—therefore represents an important conceptual outcome.

Synthetic-media literacy should teach a behavioral sequence:

Encounter → Emotional Check → Source Check → Context Check → Independent Verification → Confidence Assessment → Share / Do Not Share

This sequence introduces intentional friction into environments otherwise optimized for immediate engagement.

6. Discussion

6.1 Why Specialized Literacy May Outperform Generic Awareness

The modeled findings support a distinction between knowing that misinformation exists and knowing what to do when questionable content appears.

A generic warning such as “AI-generated videos may be fake” may increase caution but provides limited procedural guidance. Huang and Hu's experimental findings are particularly relevant because technique-based instruction produced stronger benefits than warnings alone.

General media literacy remains valuable, but contemporary synthetic media introduces additional questions:

- Is the apparent speaker actually the person shown?
- Could the voice have been cloned?
- Does the original recording exist?
- Was genuine footage altered rather than entirely generated?
- Is the surrounding caption manipulating the interpretation of authentic media?
- Does reliable provenance evidence exist?
- These questions require AI-specific literacy layered on conventional source evaluation.

6.2 Resistance Without Generalized Distrust

A major risk of aggressive deepfake-awareness campaigns is what might be called defensive overcorrection. Individuals who repeatedly hear that any photograph, recording, or video could be fabricated may respond by distrusting authentic journalism, institutions, scientific communication, or documentary evidence.

Vaccari and Chadwick's findings regarding uncertainty are particularly important in this respect. Government scientific advisers in the United Kingdom have likewise cautioned that awareness of deepfakes can contribute to broader uncertainty and institutional distrust, while the causal effect of disinformation on major behaviors should not be overstated.

Synthetic-media education should therefore promote calibrated skepticism, not blanket skepticism.

Calibrated skepticism means that users:

- demand stronger evidence for extraordinary claims;
- check sources before amplification;
- adjust confidence in proportion to available evidence;
- distinguish absence of verification from proof of fabrication; and
- remain willing to revise judgments when stronger evidence appears.

This approach preserves trust as something that can be earned through evidence rather than eliminating trust altogether.

6.3 Educational, Platform, and Policy Implications

Responsibility for synthetic-media resilience cannot reasonably be transferred entirely to individual users. Even highly literate citizens operate within platforms designed around speed, visibility, engagement, recommendation systems, and social feedback.

The UK Government Office for Science explicitly identifies the need to consider the distribution of responsibility among citizens, content providers, and governments. UNESCO similarly argues for approaches extending beyond individual detection toward broader knowledge ecosystems capable of supporting reliable collective sense-making.

An effective societal strategy should therefore integrate:

- Educational institutions, which can teach verification, AI literacy, source reasoning, and ethical content creation.
- Digital platforms, which can improve provenance signals, reporting systems, contextual information, and friction around questionable viral content.
- News organizations, which can explain verification processes and disclose uncertainty rather than merely labeling content true or false.
- Technology developers, which can strengthen provenance, watermarking, authentication, and transparent detection systems.
- Governments and civil society, which can support public education while avoiding policies that unnecessarily restrict legitimate expression.
- The objective should be an information environment in which authenticity is easier to establish and manipulation is more costly to distribute successfully.

7. Conclusion

Synthetic media represents a structural change in digital communication because realistic fabrication increasingly challenges the traditional assumption that seeing or hearing constitutes sufficient evidence. Public responses based exclusively on deepfake-detection software or visual inspection are therefore unlikely to provide comprehensive protection.This study proposed synthetic media literacy as a

multidimensional form of digital resilience integrating AI awareness, perceptual caution, source evaluation, provenance reasoning, lateral verification, emotional self-regulation, calibrated uncertainty, and responsible sharing.

Using an explicitly labeled simulated sample of 600 digital-media users, the analysis modeled three educational conditions. The synthetic-media-literacy condition produced the strongest increase in the Digital Manipulation Resistance Score, rising from 52.0 to 72.6, compared with 52.4 to 63.8 following general media literacy and 52.1 to 53.0 under the control condition.These numerical outcomes should not be interpreted as empirical population estimates. Their purpose is to demonstrate a reproducible analytical framework for future experimental testing.

The principal theoretical contribution is the argument that the success of synthetic-media education should not be evaluated solely by asking whether people can correctly label individual items as “real” or “fake.” A more meaningful measure is whether people can respond intelligently under uncertainty.Public resistance to manipulation is therefore best understood as an interaction among knowledge, verification, judgment, emotion, and behavior. The most resilient digital citizen is not the person who distrusts everything, but the person who knows when confidence is justified, when verification is necessary, and when information should not yet be shared.

Future empirical research should test the proposed model using preregistered randomized experiments, culturally diverse samples, multilingual synthetic-media stimuli, delayed follow-up assessments, and authentic social-media decision environments. Particular attention should be directed toward the durability of intervention effects because recent evidence suggests that some gains in self-efficacy and perceived literacy may weaken over time even when detection performance initially improves.

References

1. Chesney, R. M., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*(6), 1753–1820. https://doi.org/10.15779/Z38RV0D15J
2. Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society, 6*(1), 1–13. https://doi.org/10.1177/2056305120903408
3. Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review, 9*(11), 39–52. https://doi.org/10.22215/timreview/1282
4. Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons, 63*(2), 135–146. https://doi.org/10.1016/j.bushor.2019.11.006
5. Floridi, L. (2018). Artificial intelligence, deepfakes and a future of epistemic uncertainty. *Philosophy & Technology, 31*, 1–4.
6. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). DeepFakes and beyond: A survey of face manipulation and fake detection. *Information Fusion, 64*, 131–148. https://doi.org/10.1016/j.inffus.2020.06.014
7. Verdoliva, L. (2020). Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing, 14*(5), 910–932. https://doi.org/10.1109/JSTSP.2020.3002101

8. Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys, 54*(1), Article 7. Note: published 2021; exclude this one if your journal strictly requires pre-2021 references.
9. Hwang, Y., Ryu, J. Y., & Jeong, S.-H. (2020). Effects of disinformation using deepfake: The protective effect of media literacy education. *Cyberpsychology, Behavior, and Social Networking*. First published online in 2021; exclude if your cutoff is strictly before 2021.
10. Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Humanities and Social Sciences Communications, 5*, Article 65. https://doi.org/10.1057/s41599-019-0279-9
11. Guess, A. M., Nagler, J., & Tucker, J. A. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances, 5*(1), eaau4586. https://doi.org/10.1126/sciadv.aau4586
12. Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news": A typology of scholarly definitions. *Digital Journalism, 6*(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143
13. Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*. Council of Europe.
14. Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition, 6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008
15. 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
16. 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
17. 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
18. Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science, 31*(7), 770–780. https://doi.org/10.1177/0956797620939054
19. Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences, 117*(27), 15536–15545. https://doi.org/10.1073/pnas.1920498117
20. Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. *ACM Computing Surveys*. https://doi.org/10.1145/3305260