*Cognitive Computing in Sociotechnical Systems,* Vol. 205, 2026, 178–185
https://doi.org/10.54941/ahfe1007372

# Human Cognitive Processing Strategies in the Detection of AI-Generated Synthetic Media

**Yue Liu**

School of Information, Florida State University, Tallahassee, FL 32306, USA

## ABSTRACT

Deepfake technology generated by artificial intelligence (AI) is becoming increasingly realistic. This technology not only challenges people's ability to judge what they see but may also influence individual thought patterns. Current deception research generally holds that applying cognitive load to deceivers can prompt them to reveal more deception cues, without considering the cognitive impact of deception detection methods on observers. Addressing this gap, this study moves beyond the limitations of focusing solely on media content or detection mechanisms by examining how individuals perceive and interpret signs of manipulation when evaluating deepfake material, particularly their attention and cognitive processing strategies in multimedia deepfake detection tasks. Participants were asked to assess whether textual, image, and video materials are authentic or fake. The results showed that participants used different strategies in allocating perceptual and cognitive resources across the three media. Text-based materials required the longest reaction times, while image-based judgments were less accurate than the other modalities. By contrast, video materials were evaluated most quickly and showed the highest accuracy. Although video information is commonly perceived as the most challenging media format to authenticate, participants were able to determine its authenticity more quickly than in other tasks. These findings highlight the importance of considering observer cognitive load in deception detection and offer theoretical implications for integrating cognitive load theory with dual process models of judgment in HCI contexts.

**Keywords:** Deepfake detection, Cognitive load theory, Dual process theory, Information evaluation

## INTRODUCTION

The rapid development of artificial intelligence (AI) over the past few decades has given rise to a new manipulating technology. This technique of modification is known as deepfake, which creates fake material indistinguishable to the naked eye (Rana et al., 2022). The new form of cyberattack is now capable of disseminating disinformation on an unprecedented scale and with complication (Lohrmann, 2024). Deepfake technology can create misinformation by generating fake videos and images to enhance deception. It can also manipulate content and behaviors that the victim never made, leading to political interference and cyberbullying.

Received January 29, 2026; Revised April 10, 2026; Accepted April 24, 2026; Available online July 20, 2026

Since the deepfake technology emerged in 2017, there have been countless cases of deepfakes, including the 2018 Obama BuzzFeed deepfake, the 2019 David Beckham anti-malaria deepfake, and the 2020 Queen's Christmas deepfake (Homeland Security, n.d.). With the increasing sophistication of AI-driven threats, it is critical to raise awareness of these new cybercrimes. When deepfakes make it difficult to distinguish real life from manipulated content, people's trust in media communications will continue to diminish. It is difficult for people to recognize false information, particularly for manipulated content, because widely disseminated misinformation typically blends with deception and truth.

Similar to manipulated images, the deceiver does not create a fake image out of nothing. They composite different real images together with false descriptions to enhance people's trust in them. Deepfake technology increases deceivers' ability to spread deception and leads the public to unintentionally share inaccurate information. For example, on Mother's Day in the United Kingdom, Kensington Palace released the first official photo of Princess Kate with her children after her 3-month absence from public view (Harrison & Coughlan, 2024; Picheta, 2024). Initially, the public celebrated Princess Kate's recovery, but some faked details were found in the photo. The Associated Press said the image appeared to be manipulated (Melley, 2024). Likewise, AI-generated content may contain some unreasonable content. Sora, for example, can generate short 1-minute videos based on textual descriptions that produce weird behaviors (Open AI, 2024), such as a person running against the grain on a treadmill. If the public can accurately identify these illogical phenomena, they can analyze and distinguish between AI-generated deceitful content and authentic content.

Many researchers have already attempted to develop new techniques focused on detecting fake content to assist users in identifying deepfake. For example, BioID uses biometrics to investigate facial deepfake photos for authentication (BioID, 2018). Intel developed FakeCatcher to determine deepfake manipulated videos (Demir, 2022). McAfee Deepfake Detector uses a deep neural network model to identify whether audio has been generated or tampered with by AI (McAfee, 2024). However, in uncertain environments, deepfake detectors may not outperform human judgment (Abeliuk et al., 2020). As methods for creating and disseminating false content continue to evolve, reliance on a single detection system is insufficient to capture all emerging deception signals. These limitations prompt my research question: if deepfake detection is constrained by training data and accuracy is compromised, what cues do people rely on when detecting deepfakes, and how does observer cognitive load shape the cues used to distinguish deepfakes from traditional forms of online deception?

## LITERATURE REVIEW

To address how individuals evaluate cues and distinguish deepfakes from traditional forms of deception, the following literature review introduces cognitive load theory and dual process theory as the study framework for understanding intuitive and analytical judgment processes.

## Cognitive Load Theory

*Cognitive load theory* is defined as optimizing learning processes by considering the limitations of human cognitive architecture (Sweller et al., 1998). When individuals engage in complex tasks that exceed their working memory capacity, they experience high cognitive load, which may hinder the achievement of learning objectives (Sweller, 1988). It explains how information processing load generated by learning tasks affects learners' ability to process new information and construct knowledge in long-term memory (Sweller et al., 2019).

As research on deception and misinformation expanded, Vrij et al. (2006) incorporated insights from cognitive load theory into deception detection. They proposed that deception is more cognitively demanding than truth telling and that increasing cognitive load during lie detection makes it more difficult for them to maintain consistency and respond naturally, thereby increasing observable behavioural cues (Vrij et al., 2006). They actively expanded the cognitive gap between deception and truthfulness by introducing cognitively demanding interventions (Vrij et al., 2008). However, most deception studies conceptualize cognitive load as a mechanism strategically imposed on deceivers to increase the difficulty of lying. When liars experience additional cognitive pressure, their ability to maintain fluent, consistent responses deteriorates, which can improve detection accuracy. Observers primarily identify cues produced by cognitively strained liars, rather than receiving cognitive support to aid their information processing (Masip et al., 2016; Walczyk et al., 2013).

Although cognitive load theory explains how the limitations of working memory constrain information processing capacity, it does not explicitly clarify how individuals use quick, intuitive or slower, deliberate reasoning to evaluate information when constrained by these limitations, particularly the impact of different media on people's ability to discern false information. Therefore, the current study incorporated a discussion of dual-process theory to explore human deception detection behavior further.

## Dual Process Theory

*Dual process theory* provides a broad cognitive framework for understanding how people make judgments when uncertain and how systematic errors can arise (Kahneman & Frederick, 2002). Early foundations of this theory can be traced to research on heuristics and biases, which demonstrated that individuals frequently rely on simplified judgment strategies when making decisions in uncertain contexts (Tversky & Kahneman, 1974).

Building on this work, dual process theory formalized the distinction between two interacting cognitive systems. System 1 is characterized as fast, automatic, intuitive, and affect-driven, generating rapid impressions and responses with minimal cognitive effort. In contrast, System 2 is slow, effortful, and deliberative, supporting analytical reasoning and conscious evaluation (Kahneman & Frederick, 2002). Within this framework, System 1 typically produces an initial intuitive judgment, while System 2 monitors and evaluates when sufficient motivation and cognitive resources are available,

can correct or override the response. Biases and decision errors occur when System 2 fails to adequately scrutinize the outputs of System 1 or when individuals rely primarily on intuition without engaging in deeper analysis (Kahneman & Frederick, 2002).

Individual differences further shape how cognitive systems operate. Cognitive reflection ability captures the extent to which individuals can inhibit intuitive System 1 responses and engage System 2 reasoning when faced with cognitively demanding problems (Frederick, 2005). When applied to deception research, dual process theory offers a valuable framework for explaining variability in deception detection performance. Deception judgments often involve rapid impressions based on verbal, nonverbal, and contextual cues, making them especially susceptible to intuitive processing. System 1 thinkers tend to integrate a wide range of cues simultaneously, some of which may be implicitly diagnostic, enabling quick judgments that can be effective when deception is obvious or involves clear inconsistencies (Walczyk et al., 2014). In contrast, System 2 processing involves deliberate analysis of specific cues, but this focus can sometimes lead individuals to rely on stereotypical or unreliable indicators of deception (Reinhard et al., 2013).

## RESEARCH DESIGN

This research adopted cognitive load theory and dual process theory as the conceptual framework to examine how participants distribute their cognitive resources across deception detection in multimedia materials.

### Participants

There were 57 university students in the sample. The participants consisting of 24 males (42.10%) and 33 females (57.90%) were all college students from a local university taking a basic IT course on data management. Participants ranged in age from 18 to 35 years. As the primary users of digital media and possessing adequate digital media literacy necessary for tasks related to evaluation of deepfake, they were invited for the study.

### Procedure

Participants completed a Qualtrics survey online after clicking on an invitation URL. The first page showed an informed consent form, which they had to agree to before proceeding to provide demographic information. Participants carried out the baseline deepfake detection task where they judged whether content was authentic without any prompts. After completing the baseline deepfake detection task, they received deception cues training for identifying deepfake content. Then, they proceeded to evaluate manipulated digital content that included news, images, and videos. Each type of content consisted of two manipulated and two authentic material sets. Participants had to rate authenticity of each item in addition to providing their confidence rating. Their reaction time for completing each media content assessment was automatically recorded by the Qualtrics survey.

## Data Analysis

The study employed descriptive data analysis to compare participants' deepfake detection accuracy across different modalities. Results indicate that detection performance across modalities ranged from moderate to high. Video demonstrated the highest accuracy ($M$ = 0.740), followed by news ($M$ = 0.720), while images showed the lowest accuracy ($M$ = 0.680). Image-based judgments also exhibited a smaller standard deviation ($SD$ = 0.245) compared to news ($SD$ = 0.283) and video ($SD$ = 0.293).

**Table 1:** Accuracy.

| | News | Image | Video |
|---|---|---|---|
| **Mean** | 0.720 | 0.680 | 0.740 |
| **Standard deviation** | 0.283 | 0.245 | 0.293 |

The time participants spent responding to each deepfake identification task was recorded. Analysis of reaction times across modalities revealed that participants took significantly longer to respond to news items than to images and videos. The average reaction time for fake news was 33.28 seconds ($Mdn$ = 24.28), whereas real news averaged 28.81 seconds ($Mdn$ = 15.70). Reaction times for images and videos generally fell within the 12–16 second range, with a median of approximately 9 seconds.

Moreover, within the same media type, reaction times varied by authenticity. For news stimuli, false content required significantly longer response times than true content. In the image category, true images ($M$ = 15.99) took longer to process than false images ($M$ = 12.80). For videos, reaction times for fake and true content were nearly identical (12.47 seconds vs. 12.18 seconds).

**Table 2:** Response time by authenticity.

| | Fake News | Real News | Fake Image | Real Image | Fake Video | Real Video |
|---|---|---|---|---|---|---|
| **Mean** | 33.28 | 28.81 | 12.80 | 15.99 | 12.47 | 12.18 |
| **Median** | 24.28 | 15.70 | 8.84 | 9.68 | 9.33 | 9.00 |

Comparing media modalities, participants spent the longest time on news stimuli ($M$ = 31.79, $Mdn$ = 25.00), significantly less time on image stimuli ($M$ = 14.40, $Mdn$ = 10.04), and the least time on video stimuli ($M$ = 12.32, $Mdn$ = 10.01).

**Table 3:** Response time by modality.

| | News | Image | Video |
|---|---|---|---|
| **Mean** | 31.79 | 14.40 | 12.32 |
| **Median** | 25.00 | 10.04 | 10.01 |

## DISCUSSION

When participants completed deepfake detection tasks involving news articles, images, and videos, images exhibited significantly lower accuracy rates than the other two media types, indicating systematic differences across formats. This pattern was even more pronounced in participants' response times across media types. Responses to news articles were notably slower than those to images and videos, suggesting that media format itself imposes differential cognitive load on observers. Even in the absence of additional distractions in this study, participants required more deliberate processing for textual and static visual content, as reflected in longer response times and greater variability in accuracy. These findings suggest that cognitive load in deepfake detection is shaped not only by task difficulty but also by how information is structured and presented across media formats.

By contrast, video content provides richer perceptual and dynamic cues, enabling faster and more consistent judgments. This pattern indicates that intuitive processing can outperform deliberate reasoning when the informational environment offers salient and diagnostic cues. When evaluating deepfake video content, individuals can quickly and automatically assess the information without incurring significant cognitive load. However, when confronted with images, particularly news content, their processing time significantly increases, clearly indicating they engage in deep analytical activities. Overall, differences in detection accuracy across content types appear to arise from the way information presentation determines observers' cognitive processing strategies.

### Implications

This study offers several theoretical and practical implications for research on deepfake detection and human-AI interaction. First, by shifting attention from automated detection systems to the observer, the findings extend deception research by conceptualizing cognitive load as a constraint experienced by evaluators. The observed differences in accuracy and response time across media modalities suggest that deepfake detection performance is strongly shaped by how information is structured and presented, not merely by the presence of deceptive content itself.

Second, the study results provide empirical support for dual-process theory in the context of synthetic media evaluation. Videos contain dynamic and multimodal cues, enabling faster and more consistent judgments, indicating a greater reliance on intuitive processing. In contrast, news articles and static images elicited longer response times and more variable judgments, suggesting increased engagement of deliberative processing under higher cognitive load. These findings challenge the assumption that slower, more analytical processing necessarily leads to superior detection performance and demonstrate that intuitive judgments can be effective when perceptual cues are salient and diagnostically informative.

### Limitations

The first limitation of the study is that it relied on a student sample drawn from a single institution, which may limit the generalizability of the results

to more diverse populations. Especially since these students all came from a specific IT course, this group's information literacy may be higher than that of other groups.

Secondly, the study employed a descriptive analytical approach, which limits causal inference. While systematic differences were observed across media modalities, future research should use experimental manipulations of cognitive load to strengthen claims regarding underlying cognitive mechanisms.

Finally, although response time was used as a metric for cognitive effort, cognitive load was not directly manipulated or measured using physiological or subjective workload measures.

### Future Study

The study conducted as a pilot test primarily aimed to examine individual responses to detecting deepfakes across different media types. The results confirmed that cognitive load varies when identifying distinct deepfake formats. Future research should expand the participant sample and account for additional factors influencing detection ability, such as information literacy and critical thinking skills.

## CONCLUSION

Overall, individuals distinguish deepfakes from traditional online deception not by relying on a single detection strategy but by adapting their cognitive processing to the structure of the media. These results emphasize the importance of considering observer cognitive load and judgment strategies in deepfake detection and suggest that human-centered approaches remain essential for addressing emerging forms of AI-generated deception.

## ACKNOWLEDGMENT

This work was supported by the Florida State University Graduate School, the Congress of Graduate Students, the Office of the Provost, and the Office of Research through the Dissertation Research Grant.

## REFERENCES

Abeliuk, A., Benjamin, D. M., Morstatter, F., & Galstyan, A. (2020). Quantifying machine influence over human forecasters. *Scientific Reports*, *10*(1), Article 15940. https://doi.org/10.1038/s41598-020-72690-4

BioID. (2018). *Deepfake detection software*. https://www.bioid.com/deepfake-detection/

Demir, Ii. (2022, November 14). *Intel introduces real-time deepfake detector*. https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html#gs.e078ie

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Harrison, E., & Coughlan, S. (2024, March 11). *Kate photo: Princess of Wales seen after saying she edited Mother's Day picture*. BBC. https://www.bbc.com/news/uk-68534359

Homeland Security. (n.d.). *Increasing threat of DEEPFAKE identities*. https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In D. Griffin, D. Kahneman, & T. Gilovich (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Lohrmann, D. (2024). Cybersecurity, deepfakes and the human risk of AI fraud. *Government Technology Magazine*. https://www.govtech.com/security/cybersecurity-deepfakes-and-the-human-risk-of-ai-fraud

Masip, J., Blandón-Gitlin, I., Martínez, C., Herrero, C., & Ibabe, I. (2016). Strategic interviewing to detect deception: Cues to deception across repeated interviews. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01702

McAfee. (2024). *McAfee® Deepfake Detector flags AI-generated audio within seconds*. https://www.mcafee.com/ai/deepfake-detector/

Melley, B. (2024, March 11). *Why the AP retracted the first official photo of the Princess of Wales since her abdominal surgery*. https://apnews.com/article/princess-wales-kate-surgery-photo-manipulated-3863e9ac78aec420a91e4f315297c348

Open AI. (2024). *Creating video from text*. https://openai.com/index/sora/

Picheta, R. (2024, March 11). *Princess of Wales apologizes for editing Mother's Day photograph*. CNN. https://www.cnn.com/2024/03/11/uk/kate-royal-photo-apology-gbr-intl/index.html

Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. *IEEE Access*, *10*, 25494–25513. https://doi.org/10.1109/ACCESS.2022.3154404

Reinhard, M.-A., Greifeneder, R., & Scharmach, M. (2013). Unconscious processes improve lie detection. *Journal of Personality and Social Psychology*, *105*(5), 721–739. https://doi.org/10.1037/a0034352

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences*, *10*(4), 141–142. https://doi.org/10.1016/j.tics.2006.02.003

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2008). *A cognitive load approach to lie detection*. https://doi.org/10.1002/jip.82

Walczyk, J. J., Harris, L. L., Duck, T. K., & Mulay, D. (2014). A social-cognitive framework for understanding serious lies: Activation-decision-construction-action theory. *New Ideas in Psychology*, *34*, 22–36. https://doi.org/10.1016/j.newideapsych.2014.03.001

Walczyk, J. J., Igou, F. D., Dixon, L. P., & Tcholakian, T. (2013). Advancing lie detection by inducing cognitive load on liars: A review of relevant theories and techniques guided by lessons from polygraph-based approaches. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00014