Seeing is Not Believing: Helping Children Differentiate Real and AI-Generated Images

Tiffany Wilson[a], Sannyu Harris[b], Frank Wilson, III[c]

[a]Florida State University, [b]North Carolina A & T State University, [c]Florida State University

Tiffany Wilson serves as the Program Coordinator of the Counselor Education Program at Florida State University. Prior to joining the FSU faculty, Dr. Wilson worked as a high school counselor and a licensed career development coordinator in North Carolina public schools. She also worked as a therapist in various clinical settings. In addition to being a Licensed Clinical Mental Health Counselor and Licensed School Counselor, Dr. Wilson is also a National Certified Counselor, a Certified Clinical Trauma Professional, a Board Certified Tele-mental Health Professional, and a Qualified Supervisor.

Sannyu McDonald Harris serves as an Assistant Professor in the Department of Counseling at North Carolina A & T State University. Prior to joining the faculty, Dr. Harris worked in clinical mental health counseling in both group practice and hospital settings. She has experience supporting individuals and organizations through wellness initiatives, workforce development, and resilience-building practices. Dr. Harris is a Licensed Clinical Mental Health Counselor Supervisor. Her research focuses on counselor burnout, resilience, professional identity development, workforce sustainability, and employee wellness across healthcare, educational, and organizational settings.

Frank Marshall Wilson III is an IT professional, creative technologist, and lifelong learner with a strong background in computer science, project management, photography, leadership, and user support. He currently serves as an IT Manager in Human Resources at Florida State University, where he supports technology operations, onboarding, knowledge management, and digital solutions that help staff work more efficiently. With experience in higher education, technical support, team leadership, mentoring, and project development, Frank is passionate about using technology to empower people and simplify everyday processes. Originally from Winston-Salem, North Carolina, Frank brings a thoughtful, people-centered approach to his work, combining technical expertise with creativity, service, and a commitment to continuous growth.

Abstract

The rapid proliferation of generative artificial intelligence (AI) has created a digital environment in which synthetic images are increasingly difficult to distinguish from reality. This shift poses significant risks for children, who may be vulnerable to misinformation, algorithmic bias, harmful social comparison, and digital fraud. Because children regularly use social media platforms where synthetic media can spread quickly, they need age-appropriate AI image literacy skills that support critical thinking, identity development, and responsible decision-making. Helping students differentiate between real and AI-generated images requires a multifaceted approach centered on hands-on exploration, visual cue recognition, source verification, and guided reflection.

Introduction

Artificial intelligence is changing the way we create, communicate, learn, and share information. In many ways, AI is an exciting tool. It can help students brainstorm ideas, support teachers with lesson planning, improve accessibility, and open doors for creativity. However, with every powerful tool comes responsibility. One of the most important responsibilities we now have as educators, parents, and technology leaders is teaching children how to recognize AI-generated images and deepfakes.

A deepfake is a photo, video, or audio recording that has been digitally created, altered, or manipulated to look real. Years ago, many people believed that if there was a picture or video of something, it must have happened. Today, that is no longer always true. A person can appear to be in a place they have never visited. Someone's face can be placed on another person's body. A fake event can look like a real news photo. A student, teacher, celebrity, or public figure can be made to appear as if they said or did something they never actually said or did. This is why training students to identify deepfakes is not just a technology lesson. It is a life skill.

The Importance of Differentiating Real Images and AI-Generated Images

It is crucial for students to differentiate real images from AI-generated images. Students are increasingly encountering a digital environment where synthetic media is used to spread misinformation, exploit cognitive biases, and reinforce harmful societal stereotypes (Ali, DiPaola, Lee, Sindato, Kim, Blumofe, & Breazeal, 2021; Miskolczi, 2026). As students form and develop opinions and perspectives about the world around them, being misled by hyper-realistic "deepfakes" can cause long-term harm to their decision-making and worldview (Ali et al., 2021).

The development of students' ability to differentiate between authentic and AI-generated images is increasingly critical within contemporary digital environments for several interrelated reasons. First, such skills are essential in combating misinformation and disinformation, as AI-generated images can be produced at scale to support "fake news" campaigns that often disseminate more rapidly and broadly than accurate information (Ali et al., 2021; Farooq & de Vresse, 2026; Sohail, Sajjad, Zafar, Iqbal, Muhammad, & Kazim, 2025). In political contexts, synthetic imagery has already been used to advance partisan agendas and carries the potential to undermine democratic processes by presenting highly credible but fabricated evidence of events (Farooq & de Vresse, 2026; Ali et al., 2021). Additionally, AI-generated content is frequently deployed by "content farms" on social media platforms to elicit strong emotional reactions such as empathy or outrage thereby overriding critical thinking processes in favor of maximizing user engagement and advertising revenue (Miskolczi, 2026).

Second, students must be equipped to recognize algorithmic bias and the reinforcement of stereotypes embedded within AI systems (Vartiainen, Kahila, Tedre, Lopez-Pernas, & Pope, 2025). Because these models are trained on large, internet-derived datasets that often reflect historical inequities, they may reproduce and amplify dominant cultural narratives, such as depicting professionals through narrow, Westernized lenses (Vartiainen, et. al, 2025; Velásquez-Salamanca, Martín-Pascual, & Andreu-Sánchez, 2025). Prolonged exposure to such biased representations can shape students' identity development and perpetuate marginalization (Zanardi, Dedò, & Landoni, 2026).

Third, the proliferation of synthetic media heightens vulnerability to fraud and exploitation, as the accessibility of these tools lowers barriers for criminal activities, including identity theft, financial scams, and the creation of non-consensual or exploitative materials (Sohail et. al, 2025; Doornbos, 2025; Cooke, Edwards, Barkoff, & Kelly, 2025). Relatedly, AI-generated images may be misrepresented as human-created works, misleading consumers and devaluing artistic labor (Ha, Passananti, Bhaskar, Shan, Southen, Zheng, & Zhao, 2024; Cenerini, Keller, Pennazza, Santonico, & Vollero, 2025).

Fourth, students must overcome what has been described as the "Good Enough Effect," wherein visually appealing AI-generated images reduce the likelihood of deeper critical evaluation, fostering false confidence in one's ability to detect inauthentic content (Zanardi, Dedò, & Landoni, 2026; Cooke et. al, 2025). Finally, the capacity to distinguish between human- and machine-generated media is vital for maintaining academic and legal integrity, particularly as evolving copyright frameworks and educational standards continue to privilege human authorship and originality (Zanardi, Dedò, & Landoni, 2026; Cooke et. al, 2025). Collectively, these competencies prepare students to engage as informed consumers, ethical creators, and responsible citizens in an increasingly AI-mediated digital landscape (Ali et al., 2021; Doornbos, 2025).

Challenges of Detecting AI-Generated Images

Children's difficulty in distinguishing between authentic and AI-generated images is shaped by a convergence of technological sophistication and developmental factors. Advances in generative artificial intelligence, including Generative Adversarial Networks (GANs) and diffusion-based models, have enabled the production of highly realistic, aesthetically refined images that can closely mimic professional photography or digital art (Ali et al., 2021; Cooke et. al, 2025; Ha et al, 2025). Empirical findings suggest that both children and adults perform only marginally better than chance when attempting to identify such images, underscoring the extent to which human perceptual systems are challenged by these technological developments (Doornbos, 2025; Cenerini et al, 2025; Cooke et. al, 2025). Compounding this issue is the "Good Enough Effect," a cognitive tendency in which individuals cease critical evaluation once an image appears sufficiently plausible or visually appealing (Zanardi, Dedò, & Landoni, 2026). For children in particular, initial satisfaction with an image often inhibits deeper inquiry into its origin, limiting their ability to detect subtle inconsistencies or embedded biases (Zanardi, Dedò, & Landoni, 2026).

Additional cognitive and developmental factors further heighten children's vulnerability (Ali et al., 2021). Research indicates that children frequently overestimate the capabilities and reliability of artificial intelligence, often interpreting AI systems as possessing human-like reasoning rather than recognizing their dependence on probabilistic data processing (Ali et al., 2021). This tendency contributes to an over-trust in machine-generated outputs, even when those outputs may be inaccurate, biased, or entirely fabricated (Zanardi, Dedò, & Landoni, 2026). Moreover, children commonly rely on heuristics, or mental shortcuts, when evaluating digital content (Farooq & de Vresse, 2026; Miskolczi, 2026). These include affect-based judgments (e.g., trusting images that evoke a sense of realism or emotional resonance), familiarity biases that favor recognizable or comforting visual patterns, and a "machine heuristic" that assumes computational outputs are inherently objective or trustworthy (Doornbos, 2025; Farooq & de Vresse, 2026; Miskolczi, 2026). Such heuristic processing reduces the likelihood of systematic analysis and increases susceptibility to deception (Farooq & de Vresse, 2026).

A lack of technical awareness also plays a significant role. Many children are unaware of the prevalence of AI-generated media within the digital platforms they regularly engage with and possess limited understanding of how such systems are trained or how outputs are generated (Vartiainen et al, 2025; Ali et al., 2021). Without this foundational knowledge, they are less equipped to critically contextualize the media they encounter (Ali et al., 2021; Vartiainen et al, 2025). Finally, contextual cues can further obscure authenticity. The framing of images such as presenting AI-generated content as children's artwork or embedding it within familiar digital environments can shape interpretation in ways that mask synthetic origins (Cenerini et al, 2025; Jin & Yuan, 2026). Visual anomalies, including inconsistencies in lighting or object positioning, may be dismissed as creative quirks rather than recognized as indicators of algorithmic generation (Cenerini et al, 2025; Jin & Yuan, 2026). Collectively, these factors illustrate the multifaceted challenges children face in navigating an increasingly AI-mediated visual landscape and underscore the importance of targeted educational interventions to support the development of critical digital literacy skills (Cenerini et al, 2025; Jin & Yuan, 2025).

Tips to Detect AI-Generated Images

Parents play a critical role in helping children develop the skills necessary to distinguish between authentic and AI-generated images by combining explicit instruction, experiential learning, and the cultivation of critical digital literacy. One effective strategy involves teaching children to recognize common visual anomalies, or "tells," that frequently appear in AI-generated content (Miskolczi, 2026). These include anatomical inconsistencies (e.g., distorted hands, misaligned features, or unnatural teeth), violations of environmental logic (e.g., improbable scenarios or improper object use), asymmetries in facial or bodily features, and departures from physical laws such as incorrect shadows or missing reflections (Miskolczi, 2026; Jin & Yuan, 2025). Additionally, irregularities in text and texture such as unreadable lettering or unnaturally smooth, "plastic-like" skin can serve as indicators of synthetic generation (Ali et al., 2021). Beyond direct instruction, parents can engage children in interactive, game-based learning activities that encourage active inquiry and critical comparison, such as asking them to distinguish between human- and AI-created media or identify inconsistencies within "deepfake" examples (Zanardi, Dedò, & Landoni, 2026). These activities are particularly effective in fostering cognitive dissonance, prompting children to question their assumptions about authenticity (Zanardi, Dedò, & Landoni, 2026).

Equally important is providing opportunities for co-creation with AI tools, which allows children to develop a firsthand understanding of both the capabilities and limitations of these systems (Zanardi, Dedò, & Landoni, 2026). Through iterative prompting and refinement such as generating and revising AI-assisted self-portraits children can begin to recognize that AI outputs are shaped by human input and training data rather than representing infallible or objective truths (Zanardi, Dedò, & Landoni, 2026). This experiential approach also supports discussions about bias, representation, and the sociocultural influences embedded within algorithmic systems (Vartiainen et al, 2025). To further strengthen critical evaluation skills, parents should address the "Good Enough Effect" by encouraging children to move beyond initial impressions of visual appeal and engage in deeper analysis of an image's accuracy, source, and underlying assumptions (Zanardi, Dedò, & Landoni, 2026). Conversations about representation for example, questioning stereotypical depictions produced by AI can further enhance awareness of bias and promote more nuanced thinking. Finally, parents can introduce age-appropriate tools and resources, such as AI-detection indicators or content from trusted, curated platforms, to model responsible media consumption and reinforce the importance of verification (Jin & Yuan, 2025). Collectively, these strategies support the development of informed, reflective, and discerning digital citizens capable of navigating an increasingly AI-mediated visual landscape (Jin & Yuan, 2025).

AI Detection Tools

There are several free and paid tools educators and students can use to help examine suspicious photos, videos, or media. These tools should not be treated as perfect, but they can support classroom discussion, media literacy lessons, and responsible decision-making.

Table 1.

Free or Free to Start Tools

AI Detection Tool	Website	Description
TrueMedia	TrueMedia.org	can help analyze suspicious digital media, including images, videos, and audio, for signs of AI manipulation. It is especially useful for media literacy instruction and public-interest verification
DeepFake-o-Meter	https://zinc.cse.buffalo.edu/ubmdfl/deep-o-meter/	is an online platform developed through the University at Buffalo Media Forensics Lab. It allows users to analyze images, video, and audio using multiple deepfake detection models. This can be helpful for classroom demonstrations because students can see that different detection models may produce different results
Hive Moderation AI Detector	https://hivemoderation.com/	allows users to upload images, videos, or audio for AI-generated and deepfake detection. It can provide probability scores and can be useful for showing students how detection systems evaluate suspicious media

TinEye Reverse Image Search	https://tineye.com/	allows students and educators to upload or search an image to see where it has appeared online. This can help students determine whether a photo is old, reused, edited, or being presented out of context.
Google Lens / Google Reverse Image Search	https://lens.google/	can also help students trace where an image appears online, compare similar images, and check whether a photo is connected to a reliable source.
InVID / WeVerify Verification Plugin	https://www.invid-project.eu/	is a browser-based verification tool designed to help journalists and fact-checkers review images and videos from social media. Educators can use it in older student lessons about news literacy, social media verification, and misinformation.
FotoForensics	https://www.fotoforensics.com/	provides tools and tutorials for digital picture analysis, including error level analysis and metadata review. It is useful for teaching students that images can carry hidden technical information, but it should be used carefully because the results require interpretation.
Forensically	https://29a.ch/photo-forensics/#level-sweep	is a free online collection of photo forensics tools. It includes clone detection, error level analysis, metadata extraction, magnification, and noise analysis. This can be useful in classroom demonstrations where students compare original and manipulated images.
AI or Not	https://www.aiornot.com/	offers AI image checking with a free tier and paid plans. This type of tool can help students test images, but results should always be discussed as possibilities rather than final proof

Table 2

Paid or Institutional AI-Dection Tools

AI Detection Tool	Website	Description
Reality Defender	https://www.realitydefender.com/	is a paid deepfake detection platform designed for organizations, institutions, and enterprises. It can analyze manipulated or AI-generated media across formats such as image, video, and audio.

Sensity AI	https://sensity.ai/	provides deepfake detection for images, videos, and audio, with a focus on forensic analysis and institutional use. It may be more appropriate for organizations, government agencies, legal settings, or school systems needing advanced review.
GetReal Security	https://www.getrealsecurity.com/	offers multimodal deepfake and manipulated media detection across image, voice, video files, and real-time streams. This is more of an enterprise or institutional solution rather than a simple classroom tool
Hive Moderation API and Dashboard	https://hivemoderation.com/	provide more advanced AI-generated content and deepfake detection services beyond the free demo. These may be useful for larger organizations or school systems that need repeated analysis or integration into existing workflows

Conclusion

The rapid advancement of generative artificial intelligence has fundamentally altered the nature of visual information, creating a digital ecosystem in which synthetic media is increasingly indistinguishable from authentic content (Baraheem, & Nguyen, 2023; Velásquez-Salamanca, Martín-Pascual, & Andreu-Sánchez, 2025). With research demonstrating that human detection accuracy often hovers between 50% and 67%, it is evident that perceptual judgment alone is no longer sufficient to reliably identify AI-generated images (Hou, Min, Pan, & Gong, 2025; Cooke et. al, 2025). This challenge is further compounded by well-documented psychological barriers, including the "Good Enough Effect" and reliance on cognitive heuristics, which can lead both children and adults to accept visually compelling content without deeper scrutiny (Zanardi, Dedò, & Landoni, 2026; Farooq & de Vresse, 2026; Miskolczi, 2026). These tendencies highlight the urgent need to move beyond passive consumption of digital media toward more intentional and informed engagement.

Central to this effort is the development of AI literacy, which emphasizes not only the ability to identify synthetic content but also a deeper understanding of how such systems are designed, trained, and deployed (Ali et al, 2021; Zanardi, Dedò, & Landoni, 2026). Educational approaches that incorporate interactive learning experiences, such as gamified detection activities and hands-on co-creation with AI tools, have demonstrated promise in helping learners recognize both the capabilities and limitations of artificial intelligence (Ali et al, 2021; Zanardi, Dedò, & Landoni, 2026). At the same time, advances in technical detection methods and the continued expertise of trained professionals underscore that identifying AI-generated media requires a combination of technological tools and human judgment (Cenerini et al, 2025; Hou et al, 2025; Ha et al, 2024; Muthaiah, Divya, Swarnalaxmi, & Vidhyasagar, 2024). However, no single strategy is sufficient. Detection systems remain vulnerable to evolving technologies, and human evaluators are susceptible to bias and cognitive shortcuts (Kolednjak, 2024; Ha et al, 2024; Miskolczi, 2026).

Equally important is the role of transparency and trust in shaping how individuals interpret and engage with digital content. Families and educators benefit from clear labeling practices and reliance on vetted sources, which provide critical layers of protection in an increasingly complex information environment (Jin & Yuan, 2025). Ultimately, preparing students to navigate this landscape requires a comprehensive, multi-pronged approach that integrates technical safeguards, ethical standards, and intentional educational practices. By fostering critical awareness, reflective thinking, and informed skepticism, educators and families can equip the next generation to serve as responsible consumers, discerning creators, and engaged citizens in a rapidly evolving, AI-driven world (Cooke et al, 2025; Miskolczi, 2026; Zanardi, Dedò, & Landoni, 2026; Ali et al, 2021).

References

AI or Not. (n.d.). *AI detector pricing*. https://www.aiornot.com/pricing

Ali, S., DiPaola, D., Lee, I., Sindato, V., Kim, G., Blumofe, R., & Breazeal, C. (2021). Children as creators, thinkers and citizens in an AI driven future. *Computers and Education: Artificial Intelligence, 2,* 1 – 11.

Baraheem, S. & Nguyen, T. (2023). AI vs. AI: Can AI detect AI-generated images*? Journal of Imaging, 9*(1-18). https://doi.org/10.3390/jimaging9100199.

Cenerini, C., Keller, F., Pennazza, G., Santonico, M., & Vollero, L. (2025). Artistic turning test: The challenge of differentiating human and AI-generated art. *Computers in Human Behavior: Artificial Humans, 6,* 1-7. https://doi.org/10.1016/j.chbah.2025.100209

Cooke, D., Edwards, A., Barkoff, S., Kelly, K. (2025). As good as a coin toss: Human detection of AI-generated content. *Communications of the ACM, 68*(10), 100-109. https://doi.org/10.1145/3729417

Doornbos, T. (2025). *Detecting AI-generating images: What constitutes successful separation between the authentic and artificial?* (Publication No. SB3E-BT15.2024-2025.3) [Bachelor Thesis, University of Groningen]. https://gmwpublic.studenttheses.ub.rug.nl/5437/

Farooq, A. & de Vresse, C. (2026). Deciphering authenticity in the age of AI: How AI-generated disinformation images and AI detection tools influence judgements of authenticity. *AI & Society, 41*(493-504).

FotoForensics. (n.d.). *Frequently asked questions.* https://www.fotoforensics.org/faq.php?c=b&show=General

GetReal Security. (n.d.). *Deepfake detection software for enterprise.* https://www.getrealsecurity.com/

Ha, A., Passananti, J., Bhaskar, R., Shan, S., Southen, R., Zheng, H., & Zhao, B. (2024). Organic or diffused: Can we distinguish human art from AI-generated images? In Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security (CCS'24), October 14-18, 2024, Salt Lake City, UT, USA. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3658644.3670306

Hive Moderation. (n.d.). *AI-generated content detection.* https://hivemoderation.com/ai-generated-content-detection

Hou, L., Min, Y. Pan, X., & Gong, Z. (2025). Distinguishing AI-generated versus real tourism photos: Visual differences, human judgement, and deep learning detection. *Information Processing and Management, 62*, 1-14. 10.1016/j.ipm.2025.104218

InVID Project. (n.d.). *InVID verification plugin.* https://www.invid-project.eu/tools-and-services/invid-verification-plugin/

Jin, Q., & Yuan, Y. (2025). "They all look mad with each other": Understanding the needs and preferences of children and parents in AI-generated images for stories. *International Journal of Child Computer Interaction, 46*, Article 100787. https://doi.org/10.1016/j.ijcci.2025.100787

Korednjak, D. (2024). *A comparison of human authored and AI-generated picturebooks in readalongs with young learners.* [Master Thesis, University of Zagreb]. https://urn.nsk.hr/urn:nbn:hr:147:726909

Miskolczi, M. (2026). The illusion of reality: How AI-generated images (AIGIs) are fooling social media users. *Computers in Human Behavior, 176*, 1 – 15.

Muthaiah, U., Divya, A., Swarnalaxmi, T., & Vidhyasagar, B. (2024). A comparative review of AI-generated vs. real images and classification techniques. Proceedings of the Fourth International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS-2024) IEEE Xplore Part Number: CFP24UG3-ART; ISBN: 979-8-3315-2963-5.

National Security Agency, Federal Bureau of Investigation, & Cybersecurity and Infrastructure Security Agency. (2023, September). *Contextualizing deepfake threats to organizations* [Cybersecurity information sheet]. https://media.defense.gov/2023/Sep/12/2003298925/-1/-1/0/CSI-DEEPFAKE-THREATS.PDF

Reality Defender. (n.d.). *Deepfake detection.* https://www.realitydefender.com/

Sensity AI. (n.d.). *Forensic deepfake detection.* https://sensity.ai/

Sohail, S., Sajjad, S., Zafar, A., Iqbal, Z., Muhammad, Z., & Kazim, M. (2025). Deepfake image forensics for privacy protection and authenticity using deep learning. *Information, 16*(4), 270. http://doi.org/10.3390/info16040270

TinEye. (n.d.). *TinEye reverse image search.* https://tineye.com/

TrueMedia.org. (n.d.). *Identifying the authenticity and veracity of AI-generated content on social media.* https://www.truemedia.org/

University at Buffalo. (n.d.). *DeepFake-o-meter: Exposing DeepFakes.* https://www.buffalo.edu/digital-scholarship-studio-network/projects/faculty-projects/DeepFake-o-meter.html

Vartiainen, H., Kahila, J., Tedre, M., Lopez-Pernas, S., Pope, N. (2025). Enhancing children's understanding of algorithmic biases in and with text-to-image generative AI. *New Media & Society, 27*(9), 5342-5368. https://doi.org/10.1177/14614448241252820

Velásquez-Salamanca, D., Martín-Pascual, M. Á., & Andreu-Sánchez, C. (2025). Interpretation of AI-generated vs. human-made images. *Journal of Imaging, 11*(7), 227. https://doi.org/10.3390/jimaging11070227

Wagner, J. (n.d.). *Forensically, free online photo forensics tools.* https://29a.ch/photo-forensics/

Zanardi I., Dedò S., & Landoni, M. (2026). On the good-enough effect: Children reflect on their AI-generated portraits. *In Proceedings of the 24th Annual ACM Interaction Design and Children Conference* (154-167). Association for Computing Machinery. https://doi.org/10.1145/313043.3733253