Digital Skills & Jobs Platform

“Telling right from wrong”:

An exploration of deepfakes, altered information and the media authentication and critical skills we need to tell what is real and what is not

This paper is part of the Digital Skills & Jobs Platform’s 2026 Digital Briefs – a series of deep-dives on key topics in digital skills and technology – produced by some of Europe’s leading experts in the field.

The DSJP is funded by the Digital Europe Programme of the European Union. Opinions expressed do not necessarily reflect the contracting authority’s official position

Table of contents

Figures

Tables

Summary

A convincing image of an event that never happened can now be made in seconds, a familiar voice cloned from a few words, a public figure made to "say" things they never said. As the tools that generate synthetic media outpace our ability to recognise their output, how can we still tell what is true?

This deep-dive maps the full spectrum of visual manipulation, from century-old "cheapfakes" such as misleading captions and doctored photos, to deepfakes built with diffusion models. It shows why recognising synthetic media by visual inspection alone is becoming increasingly difficult, and why that makes digital literacy skills even more important.

Drawing on research and the EU's Digital Competence Framework, we set out four lines of defence: critical thinking, verification, technical detection and provenance, and regulation. Citizens need a mix of skills, from recognising synthetic content to critical thinking and concrete verification techniques. And reducing the impact of image-based disinformation takes a mix of measures, with a confident, critical citizenry at its core.

Keywords

#deepfakes, #synthetic media, #media literacy, #critical thinking, #cheapfakes, #manipulation

About the author

Irina Paraschivoiu is a researcher and entrepreneur working at the intersection of extended reality (XR), media literacy, and human–computer interaction. Over the past decade, she has designed, led, and scaled interdisciplinary R&D projects that blend technology, media, and society, turning experimental ideas into evidence-based products. As Chief Operations Officer at Polycular, she leads the development of immersive learning tools, including Escape Fake, a multi-awarded augmented reality game that strengthens resilience to disinformation and has reached over 130,000 users in seven languages. Her work spans mixed-methods evaluation, game-based learning, and design research, recognised through the European Digital Skills Award and the Ars Electronica x Culttech Award. Completing a PhD in human–computer interaction, she has published in academic venues on media and digital literacy, augmented reality, and citizen engagement.

When seeing is no longer believing

When the line between fact and fiction blurs, how can you still tell what is true? Until the past decade, most of us rarely asked this question of a photograph, a voice recording or a video. Today, a convincing image of an event that never happened can be produced in seconds, a familiar voice can be cloned from a few words of audio, and a public figure can be made to "say" things they never said. The tools that do this are improving faster than our collective ability to recognise their output, and that gap has consequences for democracy, for the workplace, and for the skills Europe will need in the years ahead.

At the centre of this shift is a family of technologies often grouped under the label **synthetic media**: images, audio and video generated or substantially altered by artificial intelligence. Synthetic media is not harmful in itself. It powers genuine creativity and is already established in film, animation and advertising. **A deepfake**, by contrast, is synthetic media with **deceptive intent**: the content is designed to pass off as real.

Yet, deepfakes are only the newest chapter in a much older story. In fact, doctored photographs and misleading captions go back well over a century. These simpler "**cheapfakes**" (Paris and Donovan 2019) remain far more common and often more damaging than their sophisticated AI cousins. Understanding that spectrum, from a misleading headline to a fully synthesised video, is the first step in building resilience.

Detecting visual manipulation has become a task for the wider public, not just for specialists; it increasingly depends on the critical-thinking, verification and media-literacy competences of ordinary citizens. Whether you are a student, a teacher, a public servant, or simply someone scrolling through a feed, the ability to ask, "could this be real, and how can I verify it?" is becoming a core digital competence (DigComp 3.0).

In this piece, we set out how to build that resilience. We explain how deepfakes are made and why they are so hard to spot; we trace the wider spectrum of visual manipulation; and we examine the lines of defence that, together, help us tell right from wrong: critical thinking, detection, provenance and the regulations emerging at European level.

Deepfakes are new, but visual manipulation is not

As the technology behind synthetic media advances, deepfakes have captured the attention of researchers, fact-checkers and policymakers because of their power to deceive and to do personal harm. Yet, visual manipulation is almost as old as photography itself. To capture the adaptation of visual manipulation techniques - from its simplest to most advanced forms - researchers Britt Paris and Joan Donovan proposed picturing them as a spectrum (Paris and Donovan 2019). At one end sit **"cheapfakes"** - manipulations that need *little or no technology*. At the other, sit **"deepfakes":** *content that requires real technical expertise and resources, chiefly machine-learning tools*.

This section walks along the cheap end of that spectrum; the next turns to the deep end.

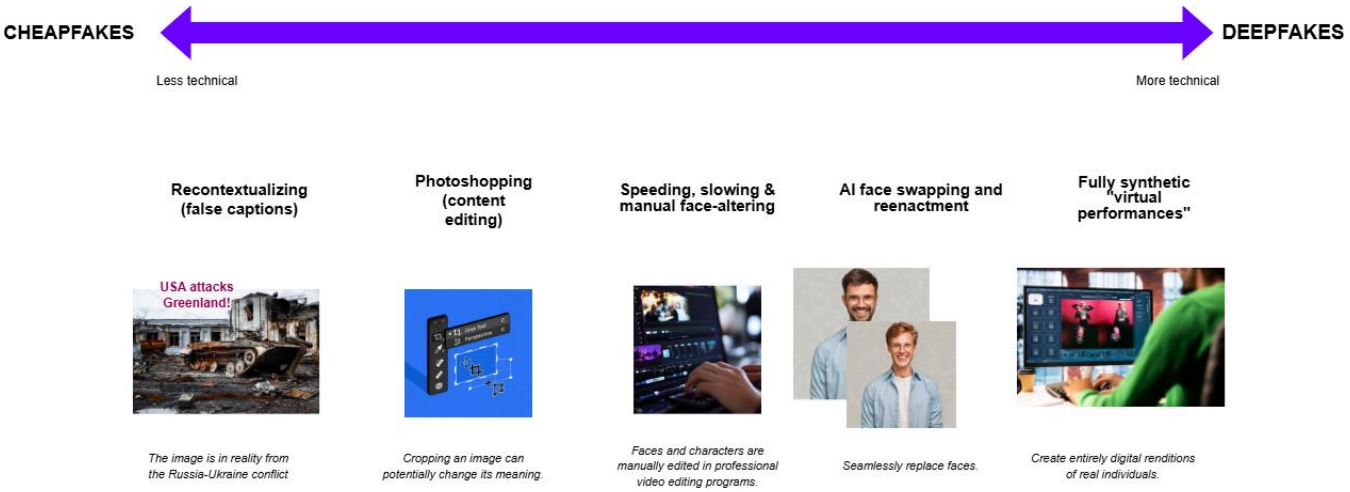

Figure 1. *The cheapfake–deepfake spectrum. Visual manipulation techniques arranged from the least to the most technically demanding. Adapted from Paris and Donovan (2019).*

Recontextualisation is the easiest form of visual manipulation and as old as photography itself. It involves false captioning: pairing a genuine, unaltered image with a misleading description, while the image itself remains untouched. The first false captioned photo dates to 1840, when French photographer Hippolyte Bayard staged a self-portrait posing as a drowned corpse (Brugioni 1999). This technique is no less potent today. In 2022, a photograph of a man and woman having fake blood applied to their faces circulated on social media as supposed proof that the war in Ukraine was being staged with "crisis actors". In reality, the image was unconnected to the war: it had been taken during the 2020 filming of a Ukrainian television series, and was debunked by the BBC, AFP and Reuters (Sardarizadeh and Robinson 2022).

A step up in effort is **content manipulation**, or "photoshopping", where the image itself is altered so that the change looks native to the scene. This covers editing text on billboards or signs, clipping and cropping, and brushing or erasing detail.

Such techniques may have become easier with image editing tools, but they long predate it: in analogue photography the same effects were achieved by physically altering negatives and prints through airbrushing, retouching and double exposure. These manipulations were used in the 1900s in several totalitarian regimes, for state propaganda. The USSR, for example, often deleted disgraced individuals from official photographs once they fell out of grace (Brugioni 1999). In one famous image, Nikolai Yezhov, a secret police official, was removed from an image in which he was walking together with Stalin.

Figure 2. *Nikolai Yezhov with Stalin along the Moscow-Volga Canal (left). Yezhov would later be removed from the image (right). The photograph was taken in 1937. Author: unknown. Public domain. Source: Wikimedia Commons*

Crude as they are, low-tech methods can even imitate deepfakes. A face can be altered or swapped by hand, frame by frame, and a video can mislead simply by being slowed, sped up or selectively cut. Everyday apps can also add basic filters and features: think of how Snapchat overlays filters on a live camera feed (Paris and Donovan 2019). The results are usually rough, but they are cheap, fast and effective. These manual face-swaps are the low-tech cousins of the AI-driven face-swapping we turn to in the next section.

While these video manipulations might be crude, cheapfakes have historically dominated and still account for most image-based misinformation as late as November 2023 (Dufour et al. 2024). AI-generated imagery has risen sharply since 2023, but in political and informational misinformation, cheapfakes still prevail. The picture is different in one disturbing area: image-based sexual abuse, overwhelmingly targeting women, where both AI-generated and "cheap" fake pornography are widespread and harmful (Ajder et al., 2019; Paris and Donovan 2019). But how are these more sophisticated manipulations produced? We will explore this in the next section.

How deepfakes are made and why detection is hard

As shown in **Figure 1**, deepfakes lie on the more sophisticated end of the visual manipulation spectrum. More narrow definitions reserve the term **deepfake** for AI-synthesized facial imagery (Rossler et al. 2019; Pei et al. 2026), whereas others take it to reference a wide range of hyper-realistic falsification of images, video and audio (Chesney and Citron 2018). Regardless of these differences, deepfakes refer to synthetic media created with machine learning models, the most common of which are **variational autoencoders** (VAEs), **generative adversarial networks** (GANs), and **diffusion models**. The process of image generation differs depending on the model used, but in the most recent years diffusion models have overtaken VAEs and GANs (Pei et al. 2026), alongside **autoregressive generation** (Liu et al. 2025). In simple terms, a diffusion model learns by adding random *"noise"* to training images until they dissolve into static, then learning to reverse the process. To create something new, it starts from pure noise and gradually *"denoises"* it into a coherent picture. An autoregressive model works differently, building an image as a sequence of

small tiles, or *"tokens"*, predicting each one from those already placed. This is the same logic that lets a large language model (LLM) predict the next word in a sentence.

Pei et al. (Pei et al. 2026) identify four main fields of deepfake research, each manipulating a face in a different way. **Face swapping** replaces one person's face with another's on existing footage. **Face reenactment** is more involved: it transfers a source person's expressions, head movements and poses onto a target face, making the target appear to move and react on command. **Talking-face generation** animates a face so that its mouth movements match a given audio clip or script: this is the technique behind convincing "lip-synced" speeches. **Facial attribute editing** changes specific features, such as expression, age or skin tone. These four categories are parallel families rather than a single ladder of sophistication, and most have shifted from GAN-based methods towards higher-quality diffusion models.

The variety of techniques for image generation helps explain why research on whether people can spot deepfakes reaches such mixed conclusions (Somoray et al. 2025). The pessimistic findings are striking. In one study, people correctly identified AI-generated faces just 48% of the time, essentially a coin toss (Nightingale and Farid 2022). Others found that people were both inclined to judge fakes as authentic and overconfident in their own judgement, a combination that leaves them especially easy to fool (Köbis et al. 2021). More optimistic studies show that accuracy can reach around 86% under favourable conditions (Groh et al. 2022). The differences in these findings partly reflect how each study was designed, the database and type of image manipulation that was used. How realistic these images were naturally impacts the accuracy with which users could label them as untruthful. But one conclusion is consistent: as the underlying models keep improving, content-based detection is increasingly difficult. That is precisely why contextual, critical and lateral thinking skills matter more than ever.

The human defence: critical thinking to identify and counter visual manipulation

Distinguishing visual manipulation is not a single skill but part of digital literacy competences. The European Digital Competence Framework, for example, integrates it under **evaluating online information** (DigComp 3.0). The skills relevant for visual manipulation detection can be grouped in four clusters: **recognising synthetic content**, **recognising manipulation and intent**, concrete **verification skills** and **building resilience in others**. Especially interesting is the complementarity between recognition of manipulation and concrete verification skills. The former looks at the broader disinformation picture: citizens should understand the role of algorithms, sources of bias and intent. The latter narrows down on content inspection of an image or a video: finding visual errors or locating the original source.

Recognising synthetic / AI-generated content	Recognising manipulation and its intent	Verification skills ("how to check")	Building resilience in others
▶ **CS1.2.03** Recognise that it can be difficult to distinguish between content generated by humans and by AI systems. ▶ **CS1.2.10** Recognise that AI systems may produce output that is inaccurate even when it seems plausible, and that the human user is responsible for checking validity. ▶ **CS1.2.08** Recognise that how an AI system is trained affects the reliability of what it produces.	▶ **CS1.2.04** Recognise examples of misinformation, disinformation, and sources of bias. ▶ **CS1.2.05** Recognise examples of social media influencing and filter bubbles. ▶ **CS1.2.12** Recognise and respond to user-directing strategies such as clickbait, nudging and gamification. ▶ **CS1.2.16** Describe the personal, social and political consequences of mis/disinformation and bias.	▶ **CS1.2.06** Make a basic assessment of the reliability and credibility of information sources and content. ▶ **CS1.2.07** Identify the source of online information and understand the purpose of fact-checking services, to develop pre-bunking and de-bunking capabilities. ▶ **CS1.2.13** Critically assess the reliability of sources, considering the role of AI, personalisation and commercial interests. ▶ **CS1.2.17** Thoroughly assess the reliability and accuracy of a diversity of sources, weighing a range of influencing factors.	▶ **CS1.2.18 / CS1.2.20** Support and help others develop the capability to critically evaluate content and build resilience to misinformation and disinformation. ▶ **CS1.2.21** Lead or contribute to initiatives supporting accurate interpretation of information online.

Table 1. Skills and competences related to the detection of visual manipulation, adaptation by author. Source: DigComp 3.0, European Digital Competence Framework, Fifth edition, Cosgrove, J., Cachia, R. Joint Research Centre, European Commission, 2025.

But how can educators, policy makers and others go about building these skills? Four types of interventions stand out as effective in the broader digital literacy domain. **Inoculation** exposes citizens to weakened doses of manipulation techniques to build psychological resistance (Basol et al. 2020; Axelsson et al. 2025). **Active learning** interventions tackle **critical thinking skills** (Paraschivoiu *et al.*, 2021; Dwyer, 2014). **Lateral reading** is inspired from fact checking strategies and is meant to train specific behaviours such as cross-referencing (Wineburg and McGrew 2019; Breakstone et al. 2021). Finally**, tips-based interventions** provide checklists and shortcuts for inspection

(Guess et al. 2020). **These interventions converge on a shared goal: to shift users from passive acceptance toward active, critical scrutiny.**

When teaching about visual manipulation, concrete and specific always beats abstract and general.

Interventions to tackle visual manipulation largely build on the four theoretical approaches. In terms of verification skills, concrete, hands-on engagement was shown to outperform interventions that simply convey information. For example, one study found that both textual and visual descriptions of most common error types in deepfake images were enough to improve detection (Geissler et al. 2026). Illogical lettering or brands, unrealistic clothing or buildings, anatomical implausibility, stylistic artifacts are cues that suggest images are not real and citizens can be trained to perform such visual inspections. Accompanying these explanations with illustrative deepfake images improved users' abilities to detect visual manipulation (ibidem).

But as the performance of diffusion models improves, visual inspection alone is no longer sufficient. **Reverse image search** and **forensic tools** that examine images for signs of manipulation remain some of the most common strategies employed by professional fact checkers. As seen above, since the proportion of cheapfakes circulating in digital environments remains high, these strategies are still robust. Forensic tools more easily point out to blurry areas, transitions that are out of place, scale or physical errors. And in the case of deepfakes, not finding any sources is a cue that an image may have been artificially generated.

In terms of recognition of manipulation, reflecting about the intent behind an image or video, whether there is any agenda or whether the narrative fits what the user knows about the respective person encourage **critical assessment**. These questions are in line with broader critical thinking literacy interventions that focus on improving analytical skills (Somoray et al. 2025). Several studies show that individuals who score high on critical thinking are more able to detect manipulation (Orhan 2023), with recommendations that media literacy should be more tightly integrated with critical thinking (Machete and Turpin 2020). This is unsurprising, since critical thinking is built on analysis, evaluation and inference, and reflective judgement: understanding the limits of knowing (Dwyer et al. 2014). These interventions are also more promising on the long run, as opposed to teaching visual inspection techniques alone, whose effects decay more quickly (Geissler et al. 2026).

Finally, showing participants an AI detector's assessment of a video also improved the accuracy of their predictions (Groh et al. 2022). This is likely because humans and machines attend to different cues, with humans processing faces holistically whereas algorithms inspect pixel-level details. That points to the fact that dealing with visual manipulation is not only a skill development issue, but one that involves human-AI collaboration, detection tools as well as regulation.

Detection, provenance and the EU response

If human detection is not enough, how can technical measures support the identification of visual manipulation? **Automated detection** relies on the same machine learning models that make the creation of synthetic media possible. For example, such models can be trained to spot inconsistencies left by manipulation or to identify "fingerprints" unique to a specific generator (Pei et al. 2026). Researchers point out that one hurdle to improving the performance of automated detection is the absence of a unified benchmark or evaluation protocol across detection methods (ibidem).

An alternative path is incorporating **watermarks** directly into the image- and video-synthesis networks themselves, a kind of digital *"fingerprint"* that enables reliable downstream identification (Nightingale and Farid 2022). A more robust approach records the origin of the material through distributed ledgers and blockchain networks, making it next to impossible to forge (Mirsky and Lee 2022). Another, proactive form of protection is to use **adversarial machine learning** defensively: adding crafted perturbations to a person's images so that deepfake networks cannot

locate or use the face. Adding *"noise"* can alter the perceived identity of the person, so that web-crawlers cannot harvest a target's images to train a model in the first place (Mirsky and Lee 2022). Finally, other researchers point to the governance of the technology itself (Nightingale and Farid 2022). Whether such powerful models should be released publicly and without restriction is an ongoing discussion that weighs the democratization of access against the dangers to democracy and wellbeing.

Governance measures require coherent policies. In the EU, several policies and guidelines regulate the creation and dissemination of synthetic media. The AI Act handles labelling at the point of creation, the Digital Services Act (DSA) handles platform-level risk and removal, whereas the Code of Practice on Disinformation (CPD) operationalises platform commitments, and media-literacy initiatives. The AI Act handles creation of synthetic media in two layers. System providers must ensure that outputs are marked in a machine-readable format and are detectable as artificially generated or manipulated. Separately, deployers such as publishers or content producers must disclose when AI is used to create realistic synthetic content. These obligations come into force on 2 August 2026.

The DSA, on the other hand, addresses the responsibilities of online platforms in which content is disseminated, such as social media. It stipulates that lawful synthetic media needs to be labelled, whereas illegal content, such as non-consensual sexually explicit deepfakes need to be removed. Platforms are also required to assess and mitigate system risks related to disinformation. The CPD further turns risk-mitigation duties into concrete commitments. It requires platforms to cooperate with fact-checkers and researchers and share data to improve detection tools.

The skills agenda for a resilient Europe

The ability to identify synthetic media is a learnable competence, not an innate talent. None of the defences we traced in the previous sections are sufficient on their own: they are complementary. Critical thinking, recognition of manipulation and verification skills build resilience against manipulations across the cheapfake-deepfake spectrum. These abilities are also an essential part of any citizen's toolbox in the 21st century. The challenge is delivery: embedding these skills from school curricula to adult education and continuing professional development. As we have shown, the development of these skills should also rely on a mix of methods. Critical thinking and inoculation build evaluation and reflective judgement skills. They are the foundation of durable resistance to manipulation. Lateral reading and fact-checking techniques instil cross-referencing habits. Concrete, hands-on practice leads to actionable insights, but needs to be refreshed, since the effects tend to decay over time.

Skills do not work in isolation. The most promising results come from pairing human judgement with technical and regulatory measures: people and detection tools attend to different cues, so combining them yields better results. As provenance standards and watermarking mature, part of the competence becomes knowing how to read and trust these signals. Resilience-building must also reach those most exposed to harm. The damage caused by synthetic media is unevenly distributed, falling hardest on women targeted by image-based sexual abuse, and any serious policy agenda must protect the most vulnerable.

As the transparency obligations of the AI Act take effect in 2026, and the DSA and CPD move from commitment to enforcement, Europe is assembling the regulatory and technical scaffolding for a healthier information space. But labels, watermarks and provenance trails only work for a public able to interpret them and willing to ask the right questions. That, ultimately, is the heart of the skills agenda for a resilient Europe: a confident and critical citizenry, harder to deceive, and better equipped to tell right from wrong.

Bibliography

Ajder, Henry, Giorgio Patrini, Francesco Cavalli, and Laurence Cullen. 2019. *Contact: Info@deeptracelabs.Com*.

Axelsson, Carl-Anton Werner, Thomas Nygren, Jon Roozenbeek, and Sander Van Der Linden. 2025. '*Bad News* in the Civics Classroom: How Serious Gameplay Fosters Teenagers' Ability to Discern Misinformation Techniques'. *Journal of Research on Technology in Education* 57 (5): 992–1018. https://doi.org/10.1080/15391523.2024.2338451.

Basol, Melisa, Jon Roozenbeek, and Sander Van Der Linden. 2020. 'Good News about Bad News: Gamified Inoculation Boosts Confidence and Cognitive Immunity Against Fake News'. *Journal of Cognition* 3 (1): 2. https://doi.org/10.5334/joc.91.

Breakstone, Joel, Mark Smith, Sam Wineburg, et al. 2021. 'Students' Civic Online Reasoning: A National Portrait'. *Educational Researcher* 50 (8): 505–15. https://doi.org/10.3102/0013189X211017495.

Brugioni, Dino A. 1999. *Photo Fakery: The History and Techniques of Photographic Deception and Manipulation*. 1. ed. Brassey's.

Chesney, Robert, and Danielle Keats Citron. 2018. 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security'. *SSRN Electronic Journal*, ahead of print. https://doi.org/10.2139/ssrn.3213954.

Dufour, Nicholas, Arkanath Pathak, Pouya Samangouei, et al. 2024. 'AMMeBa: A Large-Scale Survey and Dataset of Media-Based Misinformation In-The-Wild'. Version 2. Preprint, arXiv. https://doi.org/10.48550/ARXIV.2405.11697.

Dwyer, Christopher P., Michael J. Hogan, and Ian Stewart. 2014. 'An Integrated Critical Thinking Framework for the 21st Century'. *Thinking Skills and Creativity* 12 (June): 43–52. https://doi.org/10.1016/j.tsc.2013.12.004.

European Commission. Joint Research Centre, Judith Cosgrove, and Romina Cachia. 2025. *DigComp 3.0 European Digital Competence Framework*. EUR. Publications Office of the European Union. https://doi.org/10.2760/0001149.

Geissler, Dominique, Claire Robertson, and Stefan Feuerriegel. 2026. 'Designing Effective Digital Literacy Interventions for Boosting Deepfake Discernment'. *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, April 13, 1–45. https://doi.org/10.1145/3772318.3790428.

Groh, Matthew, Ziv Epstein, Chaz Firestone, and Rosalind Picard. 2022. 'Deepfake Detection by Human Crowds, Machines, and Machine-Informed Crowds'. *Proceedings of the National Academy of Sciences* 119 (1): e2110013119. https://doi.org/10.1073/pnas.2110013119.

Guess, Andrew M., Michael Lerner, Benjamin Lyons, et al. 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): 15536–45. https://doi.org/10.1073/pnas.1920498117.

Köbis, Nils C., Barbora Doležalová, and Ivan Soraperra. 2021. 'Fooled Twice: People Cannot Detect Deepfakes but Think They Can'. *iScience* 24 (11): 103364. https://doi.org/10.1016/j.isci.2021.103364.

Kops, Maxime, Catherine Schittenhelm, and Sebastian Wachs. 2025. 'Young People and False Information: A Scoping Review of Responses, Influential Factors, Consequences, and Prevention Programs'. *Computers in Human Behavior* 169 (August): 108650. Available at: https://doi.org/10.1016/j.chb.2025.108650.

Liu, Baoping, Bo Liu, Tianqing Zhu, and Ming Ding. 2025. 'A Review of Deepfake and Its Detection: From Generative Adversarial Networks to Diffusion Models'. *International Journal of Intelligent Systems* 2025 (1): 9987535. Available at: https://doi.org/10.1155/int/9987535.

Machete, Paul, and Marita Turpin. 2020. 'The Use of Critical Thinking to Identify Fake News: A Systematic Literature Review'. In *Responsible Design, Implementation and Use of Information and Communication Technology*, edited by Marié Hattingh, Machdel Matthee, Hanlie Smuts, Ilias Pappas, Yogesh K. Dwivedi, and Matti Mäntymäki, vol. 12067. Lecture Notes in Computer Science. Springer International Publishing. Available at: https://doi.org/10.1007/978-3-030-45002-1_20.

Mirsky, Yisroel, and Wenke Lee. 2022. 'The Creation and Detection of Deepfakes: A Survey'. *ACM Computing Surveys* 54 (1): 1–41. Available at: https://doi.org/10.1145/3425780.

Nightingale, Sophie J., and Hany Farid. 2022. 'AI-Synthesized Faces Are Indistinguishable from Real Faces and More Trustworthy'. *Proceedings of the National Academy of Sciences* 119 (8): e2120481119. Available at: https://doi.org/10.1073/pnas.2120481119.

Orhan, Ali. 2023. 'Fake News Detection on Social Media: The Predictive Role of University Students' Critical Thinking Dispositions and New Media Literacy'. *Smart Learning Environments* 10 (1): 29. Available at: https://doi.org/10.1186/s40561-023-00248-8.

Paraschivoiu, Irina, Josef Buchner, Robert Praxmarer, and Thomas Layer-Wagner. 2021. 'Escape the Fake: Development and Evaluation of an Augmented Reality Escape Room Game for Fighting Fake News'. *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play*, October 15, 320–25. Available at: https://doi.org/10.1145/3450337.3483454.

Paris, Britt, and Joan Donovan. 2019. *Deepfakes and Cheapfakes: The Manipulation of Audio and Visual Evidence*. Data and Society Research Institute.

Pei, Gan, Jiangning Zhang, Menghan Hu, et al. 2026. 'Deepfake Generation and Detection: A Benchmark and Survey'. *ACM Computing Surveys* 58 (11): 1–41. Available at: https://doi.org/10.1145/3801962.

'Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services'. 2022. CELEX: 32022R2065. European Parliament and Council. Available at: https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng.

'Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence'. 2024. CELEX: 32024R1689. European Parliament and Council. Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

Rossler, Andreas, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Niessner. 2019. 'FaceForensics++: Learning to Detect Manipulated Facial Images'. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, October, 1–11. Available at: https://doi.org/10.1109/ICCV.2019.00009.

Sardarizadeh, Shayan, and Olga Robinson. 2022. *Ukraine Invasion: False Claims the War Is a Hoax Go Viral*. March 8. Available at: https://www.bbc.com/news/60589965.

Somoray, Klaire, Dan J. Miller, and Mary Holmes. 2025. 'Human Performance in Deepfake Detection: A Systematic Review'. *Human Behavior and Emerging Technologies* 2025 (1): 1833228. Available at: https://doi.org/10.1155/hbe2/1833228.

'The Strengthened Code of Practice on Disinformation'. 2022. European Commission. Available at: https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation.

Wineburg, Sam, and Sarah McGrew. 2019. 'Lateral Reading and the Nature of Expertise: Reading Less and Learning More When Evaluating Digital Information'. *Teachers College Record: The Voice of Scholarship in Education* 121 (11): 1–40. Available at: https://doi.org/10.1177/016146811912101102.