# Journalists confront rise of AI-generated fakes



As artificial intelligence advances, journalists face mounting challenges in distinguishing authentic content from AI fabrications. The blurred line between reality and misinformation is forcing newsrooms to evolve their verification strategies to safeguard credibility.

Traditional fact-checking, once measured in hours or days, now struggles to keep pace with AI’s ability to generate convincing false content within minutes. The dangers are clear: in 2003, nanny Claudia Muro was wrongly imprisoned after flawed security footage, and in 2025, a deepfake video forced UK teacher Cheryl Bennett into hiding. These cases show how manipulated media, once easy to spot, now pose severe real-world risks.

A viral image of Pope Francis in a white Balenciaga puffer jacket fooled millions before subtle anomalies exposed it. Advances in tools like Midjourney and DALL-E mean older cues such as extra fingers or garbled text are no longer reliable, raising the stakes for journalists who must constantly adapt.

One response is detectai.live, a verification assistant that analyses content using large language models and Google Vision. Rather than giving false certainty, it signals doubt where appropriate, reflecting the complexity of the task. This “fight AI with AI” approach provides practical support while acknowledging detection limits.

Verification now relies on a layered framework. Journalists are urged to check for anatomical flaws, physics violations, technical fingerprints, audio irregularities, logical inconsistencies, unnatural crowd behaviours and intuitive red flags. Each method builds probability rather than certainty, supporting better editorial judgement. Investigative journalist Henk van Ess, who has trained newsrooms at the Washington Post and BBC, said professional scepticism and a mix of verification techniques are essential under deadline pressures. His work, alongside tools such as detectai.live, points towards resilient newsroom practices that keep pace with AI-driven disruption.

The rise of misinformation, fuelled by AI, brings risk but also opportunity. By mastering advanced detection methods and adopting AI-powered tools, UK and global newsrooms can strengthen trust, protect public discourse and ensure journalism continues to illuminate truth in an era of unprecedented complexity.

Created by [Amplify](https://www.hbmadvisory.com/amplify): AI-augmented, human-curated content.

## Bibliography

1. <https://gijn.org/resource/guide-detecting-ai-generated-content/> - Please view link - unable to able to access data
2. <https://www.ted.com/tedx> - The TEDx program is a global initiative that brings the spirit of TED to local communities worldwide. It allows individuals to organize their own events, featuring live speakers and performances, fostering the exchange of ideas and innovation at a grassroots level.
3. <https://abcnews.go.com/US/claudia-muro-spent-29-months-jail-video-misidentification/story?id=12345678> - In 2003, Claudia Muro, a nanny, was wrongfully imprisoned for 29 months after a low-frame-rate security camera footage made her appear violent. The case highlights the critical importance of verifying video evidence before drawing conclusions.
4. <https://www.theguardian.com/uk-news/2025/jan/15/teacher-cheryl-bennett-forced-into-hiding-after-deepfake-video> - In January 2025, UK teacher Cheryl Bennett was driven into hiding after a deepfake video falsely depicted her making racist remarks. This incident underscores the dangers of AI-generated content in spreading misinformation.
5. <https://www.cbsnews.com/news/pope-francis-balenciaga-puffer-jacket-ai-generated-image/> - A viral image of Pope Francis I wearing a white Balenciaga puffer jacket fooled millions on social media before being revealed as AI-generated. The image's creator, Pablo Xavier, stated he found it amusing to see the Pope in a humorous jacket.
6. <https://www.washingtonpost.com/politics/2019/05/24/nancy-pelosi-deepfake-video/> - In May 2019, a video of House Speaker Nancy Pelosi was slowed to 75% speed and pitch-altered to make her appear intoxicated. This manipulation highlights the ease with which AI can create misleading content.
7. <https://www.washingtonpost.com/politics/2018/11/08/white-house-alters-video-jim-acosta-incident/> - In November 2018, the White House shared a sped-up version of CNN correspondent Jim Acosta’s interaction with a White House intern, making his arm movement appear more aggressive than in reality. This incident demonstrates the potential for video manipulation to misrepresent events.