CENTER FOR APPLIED AI
@ TRAINING CENTERS

PLAGIARISM AND DEEPFAKES:

THE CHALLENGE OF ORIGINALITY IN THE AGE OF AI

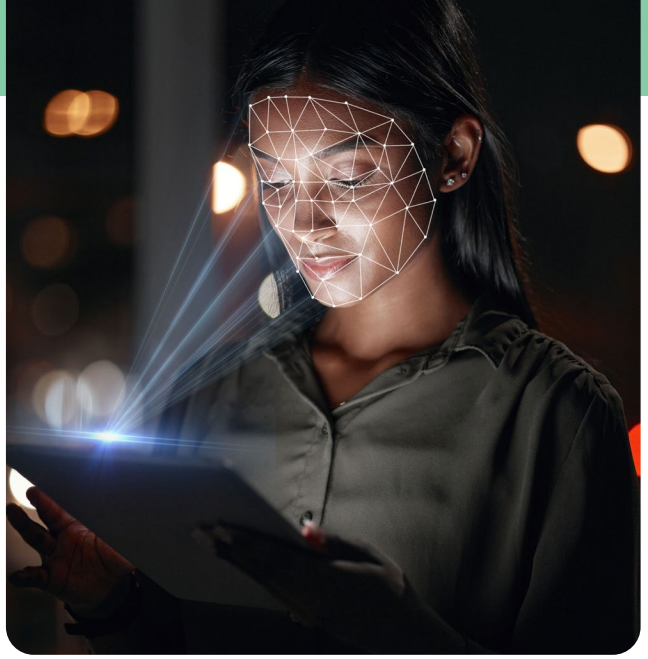
By Ommo Clark and Tom Cain

Content is often touted as being king, which simply means that good quality content is expert, authentic and trustworthy. However, these gold standards are increasingly under threat due to new, sophisticated AI tools capable of creating realistic human-like content, blurring the lines between original and fabricated. The rise of deepfakes, AI-generated media that convincingly mimic real people and events, poses a significant challenge to content integrity. As this technology becomes more accessible and refined, the need for effective detection and mitigation strategies is more critical than ever.

Maintaining the integrity of original content - defined as material that is created independently, neither replicated from existing works, nor from previously unpublished works in any identical form is paramount to foster trust, demonstrate thought leadership, and minimize the spread of inaccurate information.

Numerous AI-enabled tools are now easily accessible and allow people to generate content with clarity, speed, and at minimal cost. This convenience has blurred the lines between original, copied, fabricated and fake content across text, audio, images, and video. This poses serious risks to credibility, aids misinformation, and erodes the value of human expertise and authenticity. A growing threat in this landscape is the rise of AI-generated deepfakes, where audio, images, and videos are skillfully manipulated to falsely portray individuals saying or doing things they never did. A notable example is the viral video of a political figure appearing to give a speech that they never actually delivered.

This deepfake video spread rapidly on social media, causing widespread confusion and misinformation. In another instance, a deepfake audio clip of a CEO's voice was used to authorize a fraudulent financial transaction, highlighting the potential financial risks associated with this technology. Relatedly, the persistent issue of plagiarism which is unattributed use of someone else's work or ideas, remains a concern.



Plagiarism is not a new phenomenon, but the increasing prevalence of AI-powered language models has led to a new form of plagiarism referred to as "AI-giarism".

In one study individuals utilize AI to generate content that closely mimics human writing or modify content that they then claim as their own, thereby potentially bypassing traditional plagiarism detection methods.[1] Other studies show a correlation between the availability of AI assistants and increased plagiarism.[2] Because of this availability, there is a greater need for enhanced detection techniques and stricter ethical guidelines.

For instance, in a recent case at a renowned university, a student submitted an essay that was flagged by plagiarism detection software as being suspiciously similar to existing works. However, it was later discovered that the student had used an AI tool to paraphrase and rearrange content from various sources in an effort to evade traditional detection methods. In another case, a popular news outlet unknowingly published an AI-generated article, which had been submitted by a freelance writer who had used an AI tool to generate the entire piece, raising questions about journalistic integrity and the responsibility of publications in verifying the authenticity of their content.

1 Chan, C. K. Y. (2023). Is AI changing the rules of academic misconduct? An in-depth look at students' perceptions of 'AI-giarism'. arXiv preprint arXiv:2306.03358.

2 Smith, J., & Wilson, L. (2022). The Ethical Implications of AI in Academic Writing: A Comprehensive Analysis. Computers & Education, 168, 104197, Johnson, R., & Hernandez, M. (2022). AI-Generated Text and Academic Integrity: An Empirical Investigation, Journal of Academic Ethics, 20(2), 137-154, Nguyen, T. Q., & Patton, J. (2022), Detecting Plagiarism in the Age of AI: Challenges and Opportunities. Journal of Educational Technology Research and Development, 70(1), 123-140.

Deepfakes and plagiarism are forms of intellectual dishonesty that mislead others about true information or the source of creative work raising significant concerns about the integrity of information and the potential for deception.

Deepfakes can create false evidence of plagiarism, or conversely, exploit the "Liar's Dividend" by casting doubt on authentic content and falsely labeling it as a deepfake. Deepfakes and plagiarism differ in their methods and mediums. However, they share the following common consequences: misrepresenting originality, distorting truth, creating a hostile environment for original content creators, eroding consumer trust in information sources, and posing ethical concerns.

To address AI-giarism and Deepfakes, AI detection tools are beginning to surface that will help people differentiate between AI-generated media and original content while maintaining the quality and credibility of content in academic and professional settings. Paradoxically, AI is being used as the tool for creating sophisticated deepfakes and plagiarized content while also being used as the tool for detecting and preventing such activities. But how good are these tools at spotting what is AI generated and what is not? Do they do what they advertise? How true are their claims? In our upcoming whitepaper, we will delve deeper into this landscape, evaluating and benchmarking 35 deepfake and plagiarism detection companies and tools.

Our aim is to provide a comprehensive overview of the current state of the art, identify best practices, and highlight areas for future research and development.

By shedding light on this critical issue, we hope to contribute to the ongoing efforts to ensure a trustworthy digital landscape where truth and originality are valued and protected.

In summary, to effectively counter the threat of deepfakes, AI-assisted plagiarism as well as protect the integrity of original content, a multi-faceted approach is required. This includes ongoing technological advancements, robust ethical guidelines, and increased awareness and education surrounding AI's capabilities and limitations. Our hope is that AI is used responsibly to protect the integrity of original content and protect individuals from reputational harm.