Flawed by Design: The Leinster Principle and the Forensic Challenge of Synthetic Media

Emily Williams
Faculty of Engineering Technology
Liverpool John Moores University
Liverpool, England
e.l.williams@2022.ljmu.ac.uk

Jason Rodgers
Police Service of Scotland
Edinburgh, Scotland
0009-0003-7148-3708
j.rodgers@ljmu.ac.uk

Dr Karl Jones
Faculty of Engineering Technology
Liverpool John Moores University
Liverpool, England
K.O.Jones@ljmu.ac.uk

Dr Sebastian Chandler-Crnigoj
Faculty of Engineering Technology
Liverpool John Moores University
Liverpool, England
S.L.ChandlerCrnigoj@ljmu.ac.uk

***Abstract*—As the complexity of synthetically generated media continues to increase, it's often assumed that sufficiently advanced systems will eventually produce content that is perfect and therefore undetectable. However, by its very perception of being perfect, AI-generated media may appear anomalous when compared with real-world media. We are already in a liminal zone where anything can be denied as fake, or dismissed as synthetic, regardless of whether it is or is not [1]. As humans we believe what we see quicker than what we hear. We tend to believe something is 'fake' when it is out of context: the classic six fingers on one hand comes to mind. That's the purpose of the Leinster Principle, where the imperfections of the synthetic media mean it is harder to detect when in context. Drawing inspiration from the architectural false facades of Leinster Square Gardens in London, this study evaluates whether the introduction of such imperfections can influence automated detection systems, whether it be in the correct or indeed incorrect direction. After applying several common signal processing methods, the results indicate that these disguises can substantially influence detector behaviours, producing changes in classification outcomes of up to 100%, with ambient noise producing an absolute mean change of approximately 35%. These findings provide preliminary empirical support for the proposed principle and suggest that contextual imperfection may present an effective adversarial strategy against synthetic media detection systems.**

Keywords—Deepfake detection, audio forensics, AI, Digital forensics

I. Introduction

Advances in synthetic media generation have enabled the creation of highly convincing artificial audio, raising significant concerns for information integrity, fraud use and digital forensic investigation. As a result, a growing number of automated detection systems have been developed to identify synthetic or manipulated audio signals. However, adversarial techniques continue to evolve alongside these detection mechanisms.

This work explores a deception strategy based not in improving the intrinsic realism of synthetic signals, but by introducing imperfections that resemble those present in true recordings. Real-world audio recordings typically contain a range of environmental and processing artefacts, including background noise, microphone coloration, compression artefacts often from transcoding, and reverberation[2]. These characteristics form part of the contextual signal environment by which genuine recordings can be identified.

This study introduces the Leinster Principle, which describes how the deliberate introduction of such contextual imperfections may increase the apparent authenticity of synthetic audio and degrade the performance of detection systems.

II. The Leinster Principle

This describes a mechanism of synthetic media deception in which authenticity is recreated through the deliberate incorporation of contextual imperfections, mimicking characteristics of genuine recordings.

Rather than relying on high-fidelity signal creation, synthetic media may achieve deception by introducing artefacts that mimic the expected properties of naturally captured media. Examples include background environmental noise, dynamic range compression, equalisation effects such as telephone-style content removal, file compression artefacts and reverberation. These elements replicate the imperfections commonly present in real-world recordings and may therefore function as plausibility cues for both human perception and automated detection systems.

Whilst related to anti-forensics strategies (forensic countermeasures), the Leinster Principle differs in its operational objective. Traditional anti-forensics (forensic countermeasures) aim to remove or obscure traces of manipulation[3], however this new principle operates through the addition of contextual artefacts designed to subvert the detection process, thereby allowing synthetic media to appear consistent with the expected characteristics of genuine signals. The addition of imperfection, not the removal, may improve the undetectability of synthetic media.

III. Experimental Overview

To explore the impact of contextual imperfection on detection outcomes across several models ranging from opensource to professional level, a dataset of synthetic and real audio samples was subjected to several common signal processing transformations intended to simulate realistic recording artefacts. To respect vendor agreements and operation sensitivities, detector identities are reported in anonymised form before full report publication.

The evaluated disguise processes included:

- Addition of ambient background noise

- File Compression
- Dynamic range compression
- Equalisation filtering (Telephone band limitation)
- Reverberation
- Sample rate reduction

Detection changes were measured using Percentage Point Change from Raw (%PCFR). This unit is defined as the change in detector classification outcomes relative to the unmodified (Raw) file.

IV. Results

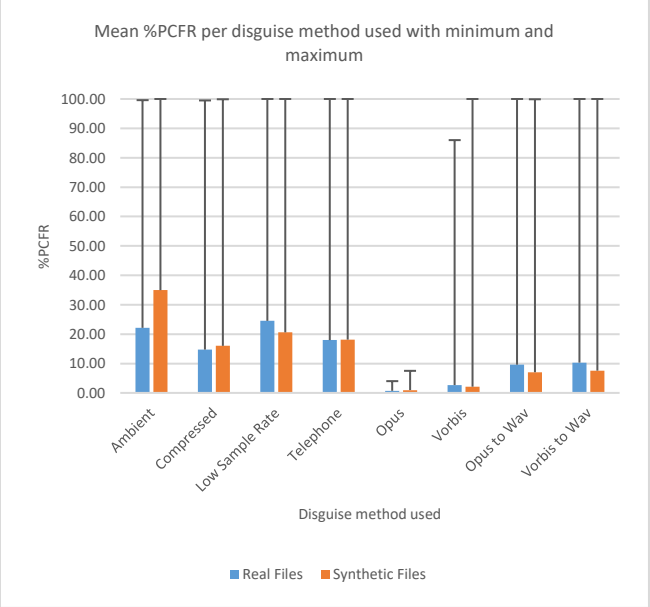
All evaluated disguise processes produced measurable degradation in detection outcomes, relative to the unprocessed baseline. Across all disguise conditions, a change was recorded for each detector.

Among the evaluated transformations, the addition of ambient noise produced the largest observed effect, with an absolute mean %PCFR of 35. Other processing techniques also produced meaningful changes in detection performance.

Some disguises in fact improved detection rates in some instances, particularly on the real-origin dataset.

Overall, each method of disguise had a registered maximum impact of 100 %PCFR. This demonstrates that each method of disguise can degrade the detection capability to such an extent that the classification itself changes – From a 'confidently real' result to a 'confidently fake' result.

These results suggest that contextual imperfection may significantly influence the behaviour of automated detection systems when evaluating audio signals.



V. Discussion and Future Work

The results provide preliminary empirical support for the Leinster Principle by demonstrating that common digital processing artefacts can measurably influence synthetic audio detection outcomes. This suggests that contextual imperfections may function as authenticity cues within both perceptual and automated evaluation processes.

The present study evaluated individual disguises in isolation, future work will explore the cumulative effects of layered disguises and additional signal manipulations to better understand how combined contextual artefacts influence detection systems.

A more comprehensive experimental analysis and formal treatment of the Leinster Principle will be presented in future work.

References

[1] B. Chesney and D. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," *Calif. Law Rev.*, vol. 107, no. 6, pp. 1753–1820, 2019, doi: 10.15779/Z38RV0D15J.

[2] H. Kazan, A. Hejase, I. Moukadem, and S. Kassem-Moussa, "Verifying the Audio Evidence to Assist Forensic Investigation," *Computer and Information Science*, vol. 14, p. 25, Mar. 2021, doi: 10.5539/cis.v14n3p25.

[3] Z. Abdullahi, N. Sagarwal, and M. Soni, "An Overview of Anti-forensic Techniques and their Impact on Digital Forensic Analysis," Mar. 2023.