Countering High-Fidelity Generative AI:
A Multi-Modal Forensic Approach

Abstract

As models like 'Nano Banana Pro' achieve photorealism, visual detection is impossible. The threat has shifted to the 'Liar's Dividend'-where real evidence is dismissed as fake. This paper proposes a Defense-in-Depth architecture integrating 'Invisible Math' (FFT), Biological Signals (rPPG), and Vision Transformers (ViT) to ensure robust verification.

1. The Problem: Optics vs. Algorithms

AI generators do not capture light; they construct pixels. This construction process leaves mathematical scars in the frequency domain that are invisible to the naked eye but detectable by algorithms.

Feature	Real Camera	AI (Nano Banana)
Sensor Noise	Unique hardware fingerprint	Gaussian/Smooth noise
Frequency	Messy, natural spectrum	Grid-like artifacts
Biology	Pulse signal (rPPG) present	No pulse / Flatline
Lighting	Consistent Physics	Inconsistent Shadows

While the eye sees a face, the algorithm sees the 'Math Scar'-grid artifacts left by upsampling layers.

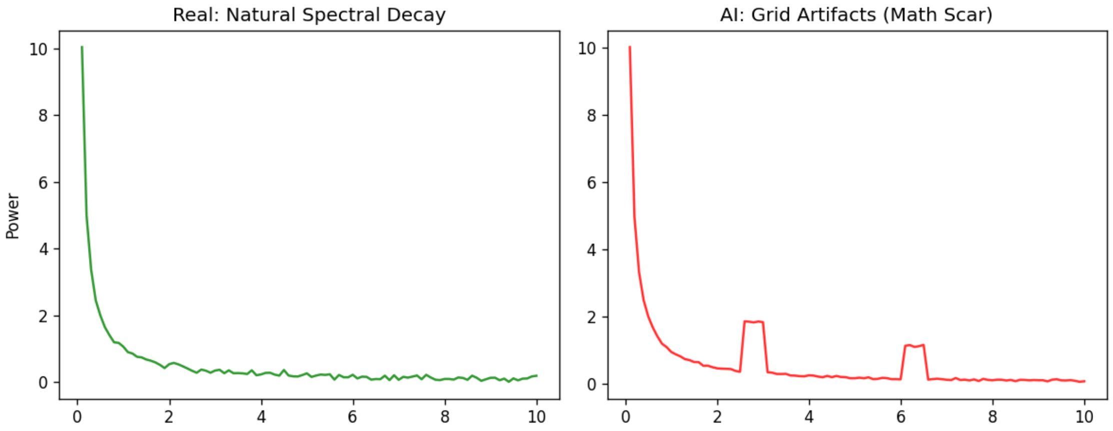


Fig 1: Real Chaos (Left) vs AI Grid Artifacts (Right)

2. Methodology: The 3-Layer Architecture

We employ a 'Defense-in-Depth' strategy. If one layer is fooled, the next catches the discrepancy.

Layer 1: Metadata & Provenance

Checks for C2PA cryptographic signatures (Adobe Content Credentials).

Layer 2: Biological Validation (rPPG)

AI generates a 'snapshot' of a face, but not the physiology. Real humans have a pulse (1.0-1.6 Hz) detectable via skin color changes. AI is biologically 'dead'.

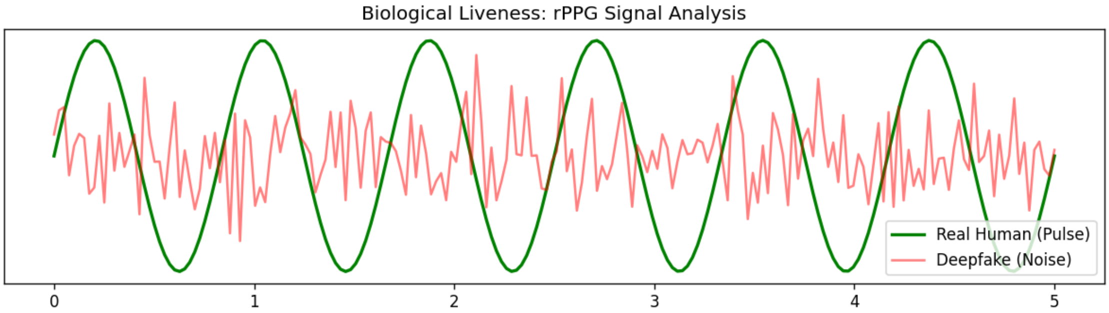


Fig 2: Real Pulse vs Deepfake Noise

Layer 3: The AI Engine

A Hybrid model combining CNNs for local texture and Transformers for global context.

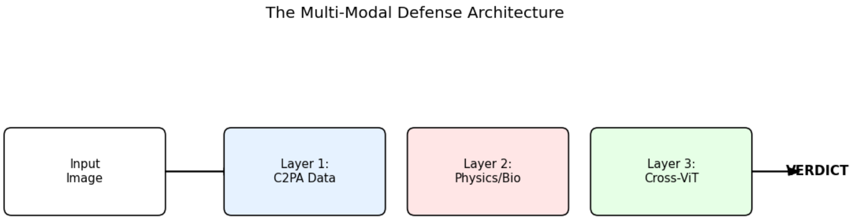


Fig 3: The System Architecture

3. Deep Technical Specifications

To defeat 'Nano Banana Pro', standard ResNets are insufficient. We utilize specialized layers designed for forensic steganalysis.

A. Pre-Processing: BayarConv

A Constrained Convolutional Layer. The center weight is fixed to -1. This acts as a high-pass filter, deleting the 'face' and leaving only the 'noise' residuals for the AI to analyze.

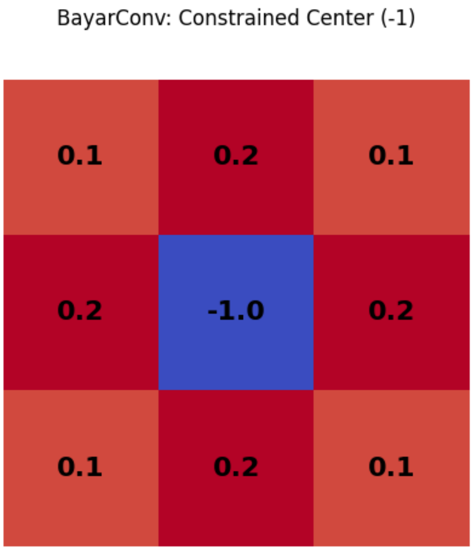
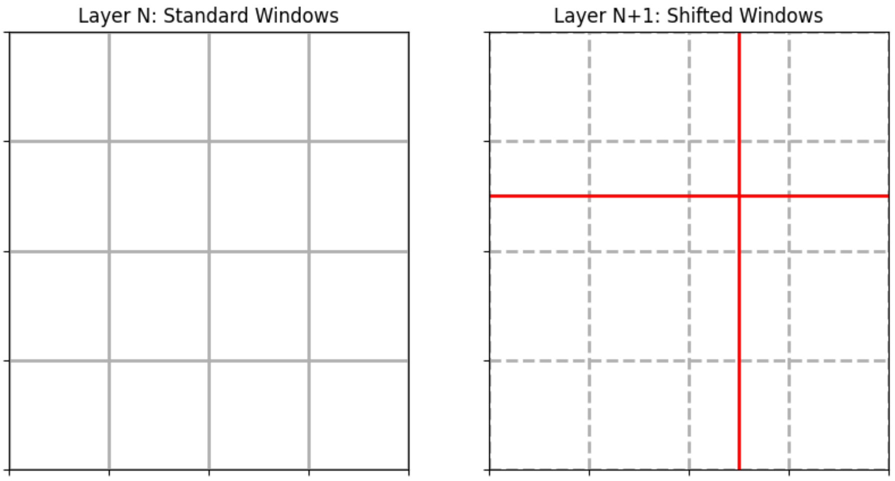

BayarConv: Constrained Center (-1)

Fig 4: BayarConv Filter Weights

B. Global Context: Swin Transformer

We use Swin Transformers with 'Shifted Windows'. This allows the model to look at relationships between different parts of the image (e.g., ear lighting vs neck shadow) to find physical inconsistencies.


Layer N: Standard Windows

Layer N+1: Shifted Windows

Fig 5: Shifted Window Attention Mechanism

4. Market Risks & Impact

The failure to detect these deepfakes results in systemic risks. Financial trading bots can be triggered by fake news, and legal proceedings can be stalled by the 'Liar's Dividend'.

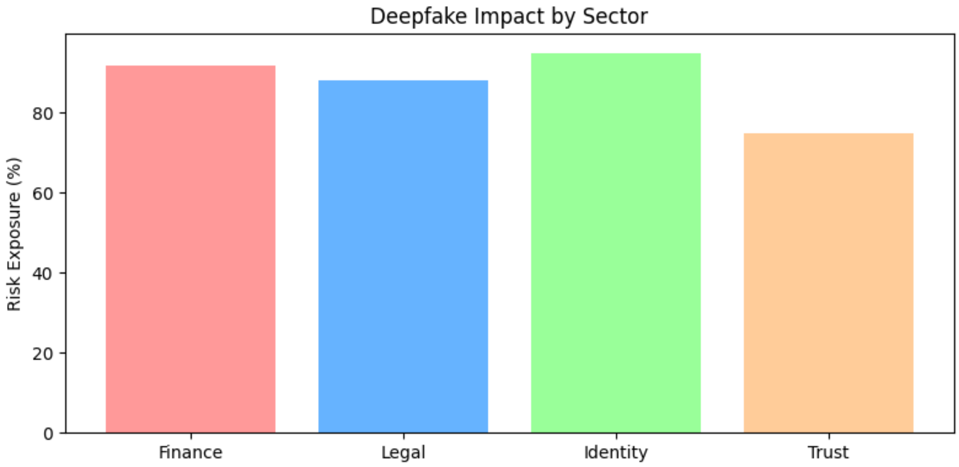


Fig 6: Risk Exposure by Sector

5. Conclusion

The era of 'seeing is believing' is over. Verification must move from Subjective (Visual) to Objective (Math/Bio). The proposed architecture provides the necessary robustness to counter next-gen Diffusion models.