DeepMinds - Abstract

Topic - Deepfake Detection & Digital Evidence Authentication System

Problem Statement

In today's digital world, discerning authentic media content from AI-generated deepfakes and manipulated visuals has become a formidable challenge. The unprecedented rise of synthetic media threatens journalistic integrity, legal proceedings, and democratic discourse, with misinformation undermining public trust and legal evidence. Manual verification is time-consuming and error-prone, while current automated tools struggle to keep pace with rapidly evolving manipulation techniques.

Mapped SDG

SDG 16: Peace, Justice, and Strong Institutions – Our solution promotes transparent information ecosystems, safeguards democratic institutions, and enhances reliable digital governance.

Existing Solutions:

Pros and Cons of Current Methods :

Tool	Target Users	Strengths	Critical Limitations
Intel FakeCatcher	Tech enterprises	96% accuracy, real-time blood flow analysis	Limited to video; not publicly accessible; requires specialized hardware
Reality Defender	Corporate/Defense	Multi-modal detection (audio/video/text)	Enterprise pricing (₹4-8L/year); black-box outputs; US-focused
Sensity AI	KYC/Identity verification	Comprehensive monitoring of 9,000+ sources	No journalist-friendly interface; lacks legal workflow integration
WeVerify/InVID	Journalists	Free browser extensions with metadata extraction	No AI deepfake detection; manual verification required; no credibility reports
Cognitech/Amped	Forensic experts	Professional-grade video analysis suites	Prohibitive cost (₹12-40L); steep learning curve; desktop-only; technical jargon outputs

Proposed Solution :

We propose a **GenAI-powered platform** that:

- **Cross-analyzes** media **formats** (***image, video, audio, text metadata***) in real time.
- Employs ***multimodal deep learning*** and *ensemble algorithms to flag manipulations.*
- Generates brief, **plain-language credibility reports** *suitable for media, legal, and public use and* ***explaining how it is fake?*** ***With proper documentation and proofs in real time.***
- **Integrates seamlessly** with *journalistic* and *legal workflows* **via APIs** and *user dashboards* .

Novelty :

- First-of-its-kind solution **combining GenAI** with *cross-format forensic analysis.*
- Delivers *intelligible credibility scores* and *manipulation evidence visualizations with* ***explanation of how we find.***
- *Supports* **vernacular language media** and **real-time scanning.**
- **Designed** for **legal & journalistic admissibility**, *scalable cloud deployment* .

Technical Description :

- Utilizes **multimodal neural networks** (*CNN, RNN, transformers*) for **cross-format media analysis.**
- Ensemble approach **combines forensic algorithms**, ***statistical watermarking***, and ***GenAI large language models.***

Input: *Digital media file (any format) + optional metadata*
Output: *Manipulation score, confidence rating, source/provenance trace, natural-language,* ***report with proper explanation and documents for explanation how it is fake.***

- **Integrates SDK/API** for **legal** and **media platforms** .

Market Potential :

Target users: Journalism (100,000+ journalists): Save *5-10 hours/week* on verification productivity boost valued at ₹50L annually per mid-size newsroom.

Legal System (3.5 crore pending cases): Reduce evidence analysis costs by **60%** through automated preliminary screening.

Government (Election Commission, PIB): Monitor campaign content for 2026-2029 state elections prevent electoral manipulation

Insurance/Banking (fraud prevention): Average deepfake fraud costs ₹3.8 crore per incident detection saves 70-80%.

Revenue Model :

Free Tier: Students, freelance journalists (5 scans/month)
Professional: ₹2,999/month (unlimited scans, API access, priority support)
Enterprise: ₹50,000/month (white-label, custom integration, SLA guar--antees)

Legal: ₹1,00,000/year (expert witness support, chain-of-custody certification)

Feasibility:

Technical: (High)

- **Pre-trained models available** (*Xception, DIRE*) ***reduce development risk.***
- **Risk**: Evolving **deepfake techniques** require continuous model retraining (*mitigated by ensemble approach + quarterly updates*).

Market: (High)

- **15 journalists** interviewed **expressed** urgent need; 3 district **courts confirmed evidence authentication** pain point.
- **Risk**: Enterprise sales cycles (6-12 months) mitigated by freemium model generating early traction.

Operational: (Modular)

- **cloud-native deployment** ensures *scalability and reliability.*

System Architecture (Workflow) :