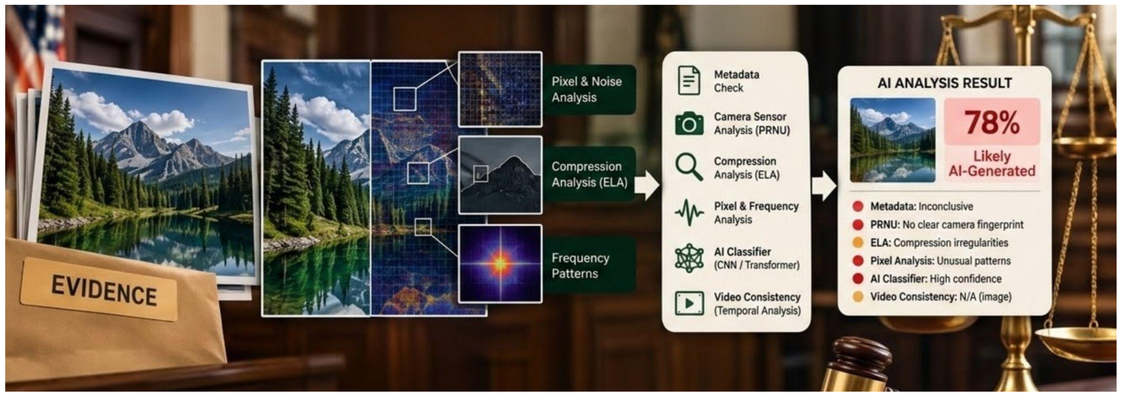
Identifying Fake Images in the Courtroom

Developing an AI-based tool to verify image evidence and protect the integrity of court proceedings.

Jillian Bogosian
Department of Computer Science
Cal Poly

Hisham Assal
Department of Computer Science
Cal Poly

CAL POLY
Noyce School of Applied Computing
COLLEGE OF ENGINEERING



Background

- AI-generated and manipulated images are becoming **easier to create and harder to identify.**
- Courtrooms increasingly rely on digital photo and video evidence, making authenticity verification more important.
- Our project explores which forensic signals can be combined to better assess whether evidence is authentic or AI-manipulated.

Research Objective

1. **Identify** the most useful methods for detecting AI-generated or manipulated images and videos.
2. Combine multiple forensic signals into **one detection workflow**.
3. Build and test a **prototype tool** that produces an interpretable confidence result.
4. Gather **feedback** from legal professionals to guide future development.

Current Tools

Current approaches for evaluating potentially AI-generated or manipulated digital evidence use a range of forensic, provenance, and AI-based methods.

Provenance & Metadata
Reviews EXIF data, timestamps, device information, Content Credentials, and known AI watermarks

Source Verification
Uses reverse image search and contextual checks to determine where media came from and whether it appeared before

Digital Forensics
Examines camera-sensor noise, compression, editing artifacts, pixel statistics, and frequency patterns

AI Detection Models
Uses machine-learning models to identify patterns associated with AI generation or manipulation

Video & Multimodal Analysis
Checks consistency across video frames, motion, audio, images, text, and surrounding context

Our goal is to integrate these methods into an **expandable framework** that can incorporate new detection techniques as they emerge.

Acknowledgments

This research was funded by the Noyce College of Applied Computing at Cal Poly, San Luis Obispo.

For any questions, comments, or a full list of resources please reach out at jbogosia@calpoly.edu.

Professional Feedback

We surveyed 26 legal professionals—including District Attorneys, Deputy District Attorneys, cybercrime prosecutors, DA investigators, and digital forensics managers across California.

26
Legal Professionals Surveyed
District Attorneys, Deputy District Attorneys, cybercrime prosecutors, DA investigators, and digital forensics professionals across California.

Experience: 7 to 32+ years

90%+
Handle Digital Evidence Regularly
of respondents handle photo or video evidence daily or in nearly every case.

Common sources include:
Body-worn cameras · Dashcams · Surveillance footage · Cell phones · Social media

Few
Use Specialized AI-Detection Tools
Authentication remains largely human-driven. Most offices do not currently use AI-detection software.

Common authentication methods:
- Witness testimony
- Chain of custody
- Hash values
- Internal forensic or cybercrime units

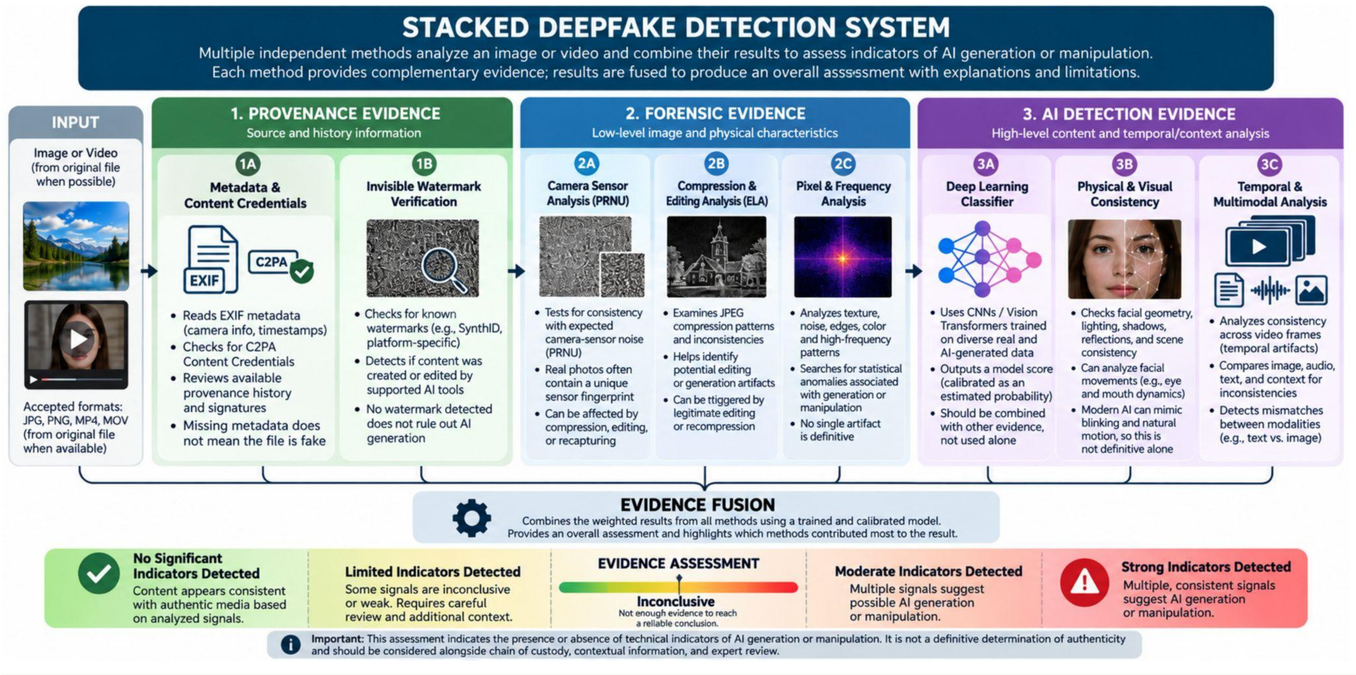
12 of 26
Interested in Follow-Up Collaboration
Nearly half of respondents volunteered for follow-up discussions, interviews, or prototype feedback.

This provides an opportunity to test and improve the system with real legal and forensic users.

Our Stack

- **Multiple detection** methods analyze different forensic signals.
- Their outputs are combined into a **single confidence score.**
- This improves interpretability compared with relying on one detector alone.

"Very soon, people will only put as much trust into pictures and video as they do into writing: not much." response from our questionnaire form



Next Steps

- Present at the 2026 Elevator Pitch Competition through Cal Poly Center for Entrepreneurship and Innovation
- Pilot the prototype with the 12 respondents from our survey
- Introduce technology as evidence in actual cases get it upheld under Kelly/Frye standard.