Spellbreaker

# Spellbreaker.ai Redesign

SI 699: Mastery Course

Project Duration

Jan 2026 - Apr 2026

Prepared by

Natasha Girish, Nikita Mahuli,
Saanch Kottige, Sandhya Sharma

# Table of Contents

# Executive Summary

## Trust is constructed through process, provenance, and defensibility.

As AI-generated media becomes increasingly sophisticated, legal professionals face a critical challenge: determining whether digital evidence is authentic. Spellbreaker, developed by ProbeTruth, uses advanced NeuroSymbolic AI to detect deepfakes across video, audio, and images. However, its reports were too technical for the attorneys, investigators, and court staff who rely on them to make high-stakes decisions.

This research investigates how legal, forensic, and adjacent professionals evaluate digital evidence in the age of deepfakes, and what Spellbreaker must provide to be credible and defensible in high-stakes contexts.

Using digital ethnography, SME interviews, and potential end-user interviews, we identified convergent patterns around trust, provenance, explainability, workflow integration, and role-based needs.

| METHODS | FINDINGS | EVALUATIONS | REDESIGNS |
|---|---|---|---|
| 5 | 8 | 4 | 4 |

# Introduction

## Motivation & Project Goals

1

### MOTIVATION

The deepfake threat is real and growing. The generative AI market is projected to grow 560% between 2025 and 2031, reaching $442B. It is projected that generative AI could drive US fraud losses from $12.3B in 2023 to $40B by 2027.

2

### THE PROBLEM

Spellbreaker produces sophisticated evidence-backed reports, but they are too technical. The people who need to act on those reports, attorneys, investigators, and court staff, cannot translate them into confident decisions.

3

### GOALS

Understand, define, and redesign. Understand how legal professionals evaluate digital evidence. Identify what makes a deepfake report credible. Define workflow requirements for Spellbreaker's reporting experience.

4

### HMW

How might we redesign Spellbreaker's end-to-end reporting experience so legal professionals can understand whether media is authentic, manipulated, or inconclusive, and why?

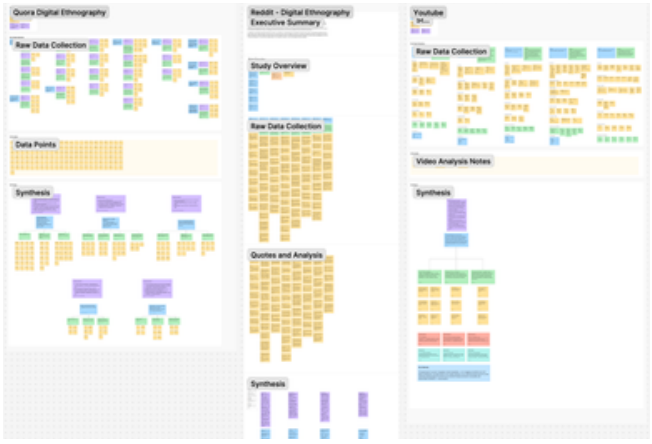
# Research

We employed a comprehensive mixed-methods research approach conducted over four weeks, combining primary qualitative research with secondary analysis to build a comprehensive understanding of how legal professionals reason about digital evidence and what Spellbreaker must provide to earn their trust.

## Approach / Research plan

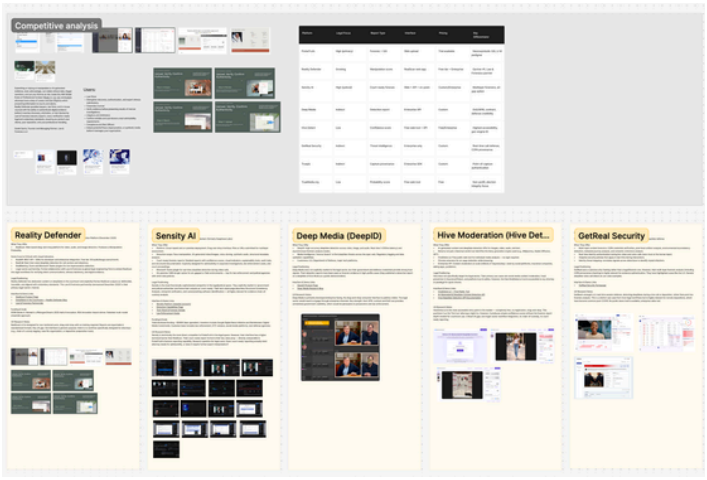
### DIGITAL ETHNOGRAPHY

Quora (2018–2026), Reddit (2020–2024), YouTube — analyzing how legal professionals reason about deepfakes and digital evidence in public discourse. Focused on identifying patterns in trust, skepticism, and decision-making across real-world scenarios.




### EXPERT & PROXY USER INTERVIEWS

7 interviews: CTO, Backend Engineer, Law Professor, Law Student, CMS Vendor, AAA Stakeholders, G1 Tech Japan. Semi-structured, 30–60 min each. Uncovered cross-functional perspectives on system behavior, legal expectations, and gaps in translating outputs into decisions.

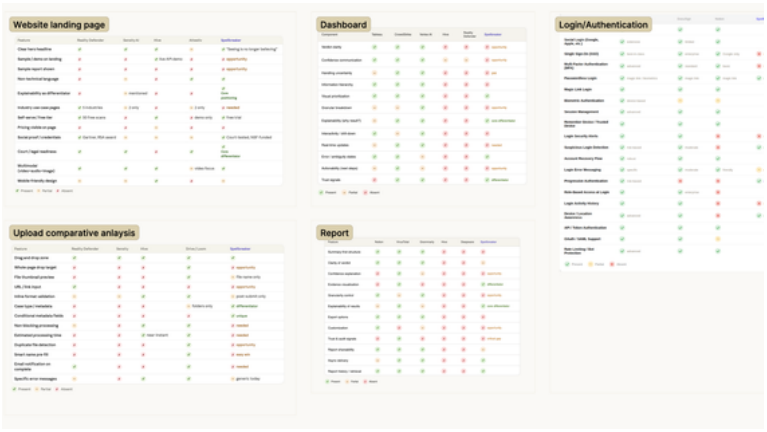



## COMPETITIVE ANALYSIS

8 platforms reviewed: Reality Defender, Sensity AI, Deep Media, Hive Detect, GetReal, Truepic, TrueMedia.org — across report quality, legal focus, and UX.

## FEATURE DESIGN BENCHMARKING

Structured evaluation of best practices for landing pages, login flows, media upload, dashboards, and report presentation across both direct competitors and enterprise SaaS platforms.




## LITERATURE REVIEW

How AI-generated findings should be communicated so legal users can understand, trust, and use them responsibly within evidentiary workflows.

# Key Findings

## 01. Courts trust process, not media

Legal professionals do not treat digital evidence as self-authenticating. They rely on provenance, chain of custody, and other metadata to decide whether a result is credible enough to act on. This means Spellbreaker must support evidentiary process, not just output a verdict.

> *Digital media does not authenticate itself. It requires a witness, a chain of custody, a party willing to stand behind it.*

## 02. Explainability is mandatory

Users do not trust unexplained confidence scores or highly technical outputs. They need clear reasoning, plain-language explanation, and visible evidence supporting the conclusion.

> *A confidence number without a clear explanation... is meaningless in a legal context.*

## 03. Role-based needs diverge

Different stakeholders need different levels of detail. Judges and arbitrators need plain-language summaries, while attorneys and experts need stronger supporting evidence and forensic depth.

> *The report is intimidating... many terms were foreign to me.*

# Key Findings

## 04. Provenance and file forensics matter more than visual inspection

Trust is tied more closely to metadata, file structure, and source history than to whether a file looks real on the surface. Reports should explain what was analyzed and why those signals matter.

> *Metadata and file traces are treated as stronger evidence than appearance alone.*

## 05. Digital evidence is contextual

Media is rarely judged in isolation. Legal users interpret authenticity in relation to surrounding evidence, case context, and corroborating details.

> *Video is one part of a broader evidence bundle.*

## 06. Deepfakes increase scrutiny; they do not collapse the system

Participants did not describe manipulated media as making digital evidence unusable. Instead, they expected more verification, more documentation, and more expert review.

> *Deepfakes raise the burden of explanation, verification, and challenge.*

# Key Findings

## 07. Trust is shifting toward infrastructure and capture integrity

Users increasingly expect authenticity systems to include provenance records, capture-time signals, and audit trails. Spellbreaker should be positioned as part of that larger workflow.

> *Trust may increasingly shift to systems that preserve provenance from the moment of capture.*

## 08. The broader climate is shifting from trust to skepticism

Deepfakes are creating a wider cultural shift in how people interpret visual evidence. In that environment, products need calm language, visible limitations, and conservative claims.

> *Wake-up call. Under attack." "Existential threat."*

Across our research, the same pattern emerged repeatedly: legal users do not need more technical detail by default. They need a reporting experience that makes verdicts easier to understand, explanations easier to follow, and supporting evidence easier to defend. Trust in this context is built through process, provenance, and clear communication, which directly shaped the design requirements for the redesigned Spellbreaker experience.

# Design Requirements

Our research findings translated into a clear set of product requirements for the redesigned Spellbreaker experience. These requirements focused on helping legal and investigative users understand digital media manipulation verdicts more easily, trust them more appropriately, and act on them more confidently within high-stakes workflows. The redesign needed to support defensibility and decision-making across the full reporting experience, rather than prioritizing technical output alone.

## Report

The report must lead with a clear verdict, explain confidence in plain language, and make supporting evidence easier to interpret. It must also note provenance and limitations with enough clarity to support legal review and challenge. Since users have different levels of expertise, the system must support layered reporting rather than a single one-size-fits-all format.

## Workflow

The experience must make key stages of the workflow easier to understand, including login, upload, processing, report review, and next steps. Upload constraints and processing expectations should be visible before users encounter problems. Additionally, the dashboard must make report retrieval, scanning, and evidence management faster and more intuitive.

## Experience

The interface must reduce cognitive load for non-technical legal users while preserving deeper technical detail for expert review. Language should remain calm, clear, and institutionally credible, avoiding overconfident claims. Across the product, the most important information should be easy to find first, with additional detail available when needed.

# Design

## Concept Overview

The redesign covers four core flows: Login & Authentication, Upload, Dashboard, and Report. Each concept was explored through multiple design directions before a final option was selected based on team preference, client feedback, and preference testing results. The design philosophy across all flows prioritized clarity over complexity, trust over accuracy claims, and decision support over raw data display.

### LOGIN

Three directions were explored before a split layout was selected. It pairs product storytelling on the left with a clean login form on the right, balancing clarity for new users and quick access for returning users. Dark imagery and the tagline set a high-stakes, forensic tone.

### DASHBOARD

Testing revealed that the original metrics were not useful. The redesign introduces a unified, table-based analysis history with thumbnails, status indicators, and key metadata supporting quick scanning and comparison.

### UPLOAD

Previously, constraints appeared only after errors. The new flow surfaces file limits upfront, organizes formats clearly, and adds in-platform trimming / compression to resolve issues without leaving the system.

### REPORT

Two report formats: basic and detailed, serve different user needs. Basic leads with verdict, confidence, and plain-language summary. Detailed provides forensic anomaly timelines, signal breakdowns, and model-level explanations. Both include QR code for verification.

# Evaluative Research

We evaluated the redesigned Spellbreaker experience through a multi-method process focused on usability, credibility, and workflow understanding. Building on earlier formative research, this phase examined whether users could move through the system, interpret results correctly, and feel confident using the platform in high-stakes legal and investigative contexts.

## Approach / Research plan

| | |
|---|---|
| **HEURISTIC EVALUATION** | Login/sign-up, upload, dashboard, and report review. Evaluating the redesigned workflow against usability heuristics to identify issues in clarity, feedback, terminology, and error recovery. |
| **INTERNAL STAKEHOLDER REVIEW** | Low-fidelity design review with internal Spellbreaker stakeholders. Assessing workflow direction, branding consistency, feature feasibility, and alignment with current product expectations. |
| **PREFERENCE TESTING** | Lyssna study with 15 participants, primarily UX designers and UX researchers. Comparing login, dashboard, and report design options to understand which layouts felt clearest, most credible, and easiest to understand. |
| **USABILITY TESTING** | Moderated sessions with 3 law students serving as proxy users. Testing login, upload, processing, report review, confidence interpretation, and supporting evidence review across the end-to-end workflow. |

# Design Evaluation

The evaluation confirmed that the redesigned workflow was understandable and usable. The most important refinements were not about changing the structure of the experience, but about helping users interpret evidence more clearly, understand limitations more easily, and act on results with greater confidence. These insights directly shaped the final refinements to the login, upload, dashboard, and reporting experience.

# Insights

Design exploration and evaluation helped refine the redesigned Spellbreaker experience from a strong overall direction into a more focused and credible workflow. Across our evalutations, one pattern remained consistent: users were generally able to move through the system successfully, but the aspects of the design that shaped trust, interpretation, and confidence needed stronger support.

## 01. Clarity and structure improved understanding

Users responded best to interfaces with strong hierarchy, clear verdict visibility, and simpler organization. In preference testing, participants consistently favored layouts that felt easier to scan and more straightforward to understand. In usability testing, participants were able to identify the main verdict quickly when the report identified the outcome and summary early.

## 02. Confidence needed explanation

One of the clearest findings from the evaluation was that confidence information was easy to locate, but not always easy to interpret. Some participants read the score as certainty rather than as a result that still required judgment. This reinforced the need for clearer explanation, stronger framing, and more accessible language around what confidence means within the report.

## 03. System feedback shaped trust across the workflow

The evaluation also showed that upload and processing states played an important role in how trustworthy the experience felt. Participants wanted clearer information about what the system was doing, what would happen next, and whether they should wait or leave. Heuristic review surfaced similar issues around system status, pending states, and error recovery, showing that transparency during these cases was essential to raise user confidence.

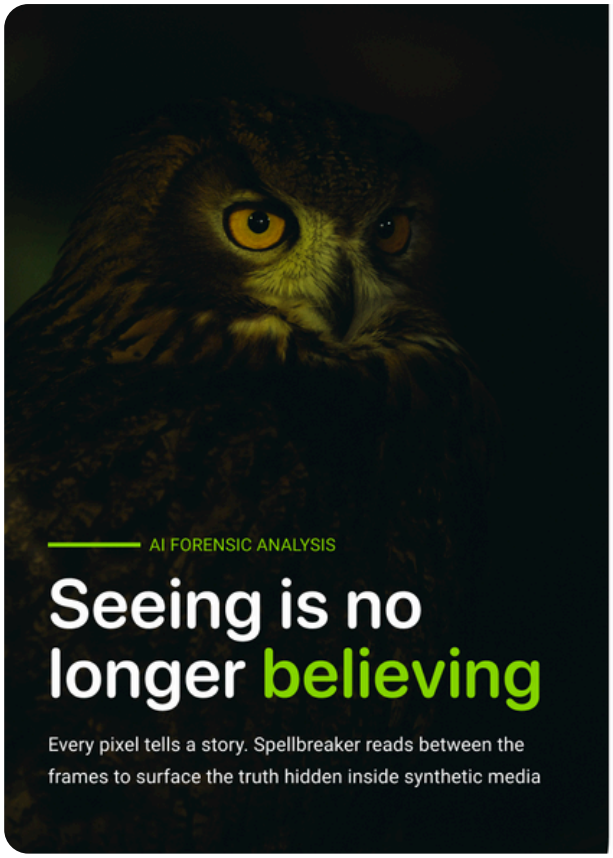
## 04. Professional presentation increased credibility

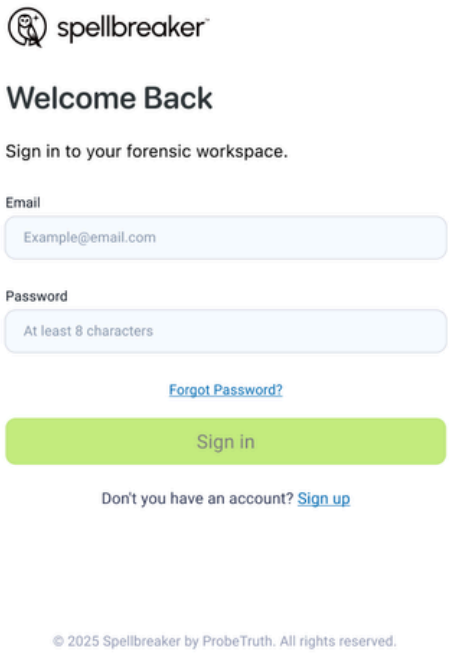
Across methods, users responded positively to cleaner, more formal layouts that felt appropriate for a legal and investigative product. Structured content, professional wording, and visually restrained presentation made the system feel more credible and easier to trust. Preference testing showed that users favored report designs with stronger hierarchy and more professional language, while usability testing confirmed that participants valued the report's structured and professional presentation of findings.

Prototype

# Final Design

## Login / Sign-up



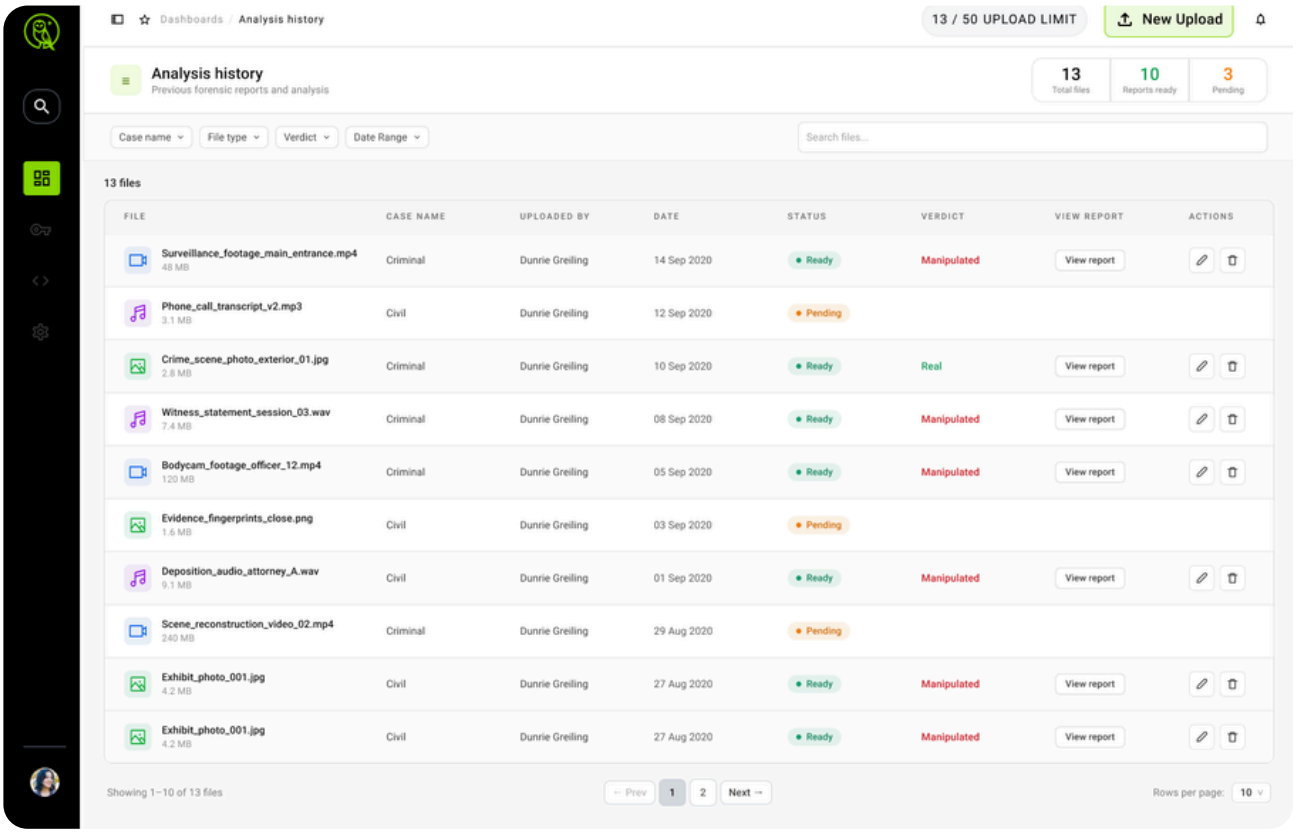




The login experience was redesigned to feel more credible and aligned with Spellbreaker's purpose. The earlier interface was functional, but it offered little context about the platform or its role in a high-stakes forensic workflow. The new split-layout design adds product context while keeping the form simple and familiar.

The visual panel reinforces the platform's forensic focus, while the sign-in form remains clean and easy to scan. This creates a more trustworthy entry point without adding friction. Preference testing and usability testing supported this direction, showing that users preferred the clearer layout and responded well to stronger context at entry.

# Final Design

## Dashboard

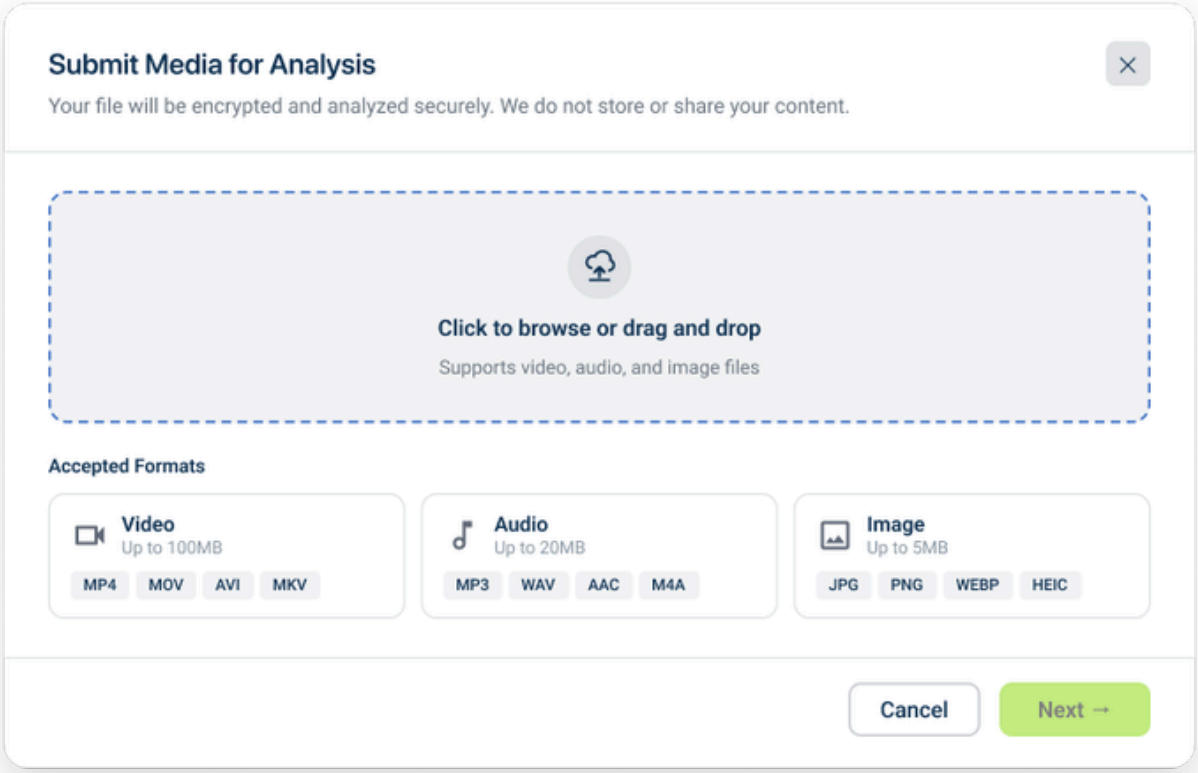



The dashboard was redesigned to shift Spellbreaker from a fragmented activity view into a clearer evidence management workspace. In the earlier version, recent activity, summary cards, and analysis history competed for attention, making it harder for users to focus on the files and reports that mattered most. The new dashboard centers the experience around a unified analysis history table that brings verdicts, statuses, reports, and actions into one place.

Search, filtering, thumbnail previews, and color-coded status indicators support faster scanning and retrieval, while role-based views reduce unnecessary complexity. This makes the dashboard more structured, easier to navigate, and better aligned with how users manage evidence in high-stakes workflows.

# Final Design

## Upload

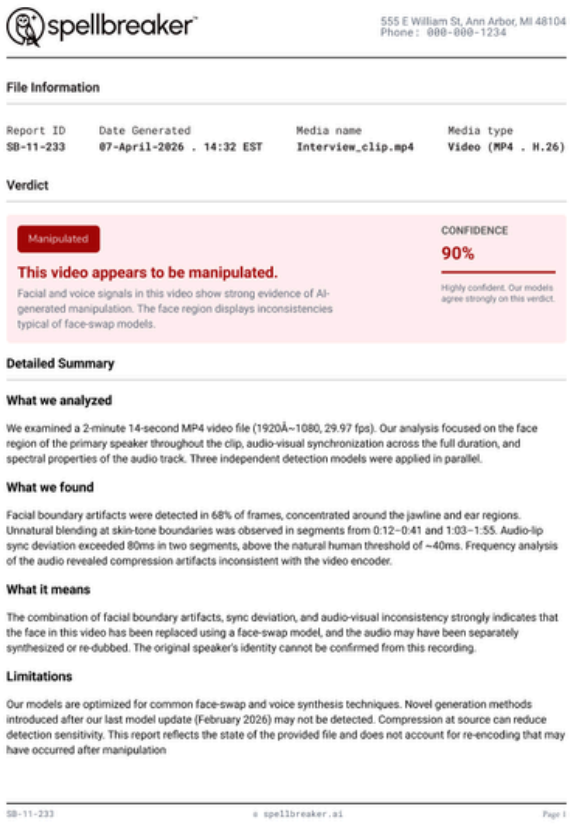



The upload flow was redesigned to be clearer and more structured. In the earlier version, upload and analysis felt compressed into one step, making it harder for users to understand what would happen next. The new flow identifies supported formats, file constraints, and analysis options earlier, reducing uncertainty before submission.

The redesigned experience also gives users more control by guiding them through file selection, case details, and analysis type, including the choice between basic and detailed reports. This makes the upload process feel more transparent and better suited to high-stakes evidence review.

# Final Design

## Report



spellbreaker

555 E William St, Ann Arbor, MI 48104
Phone: 000-000-1234

**File Information**

| Report ID | Date Generated | Media name | Media type |
|---|---|---|---|
| SB-11-233 | 07-April-2026 . 14:32 EST | Interview_clip.mp4 | Video (MP4 . H.26) |

**Verdict**

Manipulated

**This video appears to be manipulated.**

Facial and voice signals in this video show strong evidence of AI-generated manipulation. The face region displays inconsistencies typical of face-swap models.

CONFIDENCE
90%
Highly confident. Our models agree strongly on this verdict.

**Detailed Summary**

**What we analyzed**

We examined a 2-minute 14-second MP4 video file (1920Â~1080, 29.97 fps). Our analysis focused on the face region of the primary speaker throughout the clip, audio-visual synchronization across the full duration, and spectral properties of the audio track. Three independent detection models were applied in parallel.

**What we found**

Facial boundary artifacts were detected in 68% of frames, concentrated around the jawline and ear regions. Unnatural blending at skin-tone boundaries was observed in segments from 0:12–0:41 and 1:03–1:55. Audio-lip sync deviation exceeded 80ms in two segments, above the natural human threshold of ~40ms. Frequency analysis of the audio revealed compression artifacts inconsistent with the video encoder.

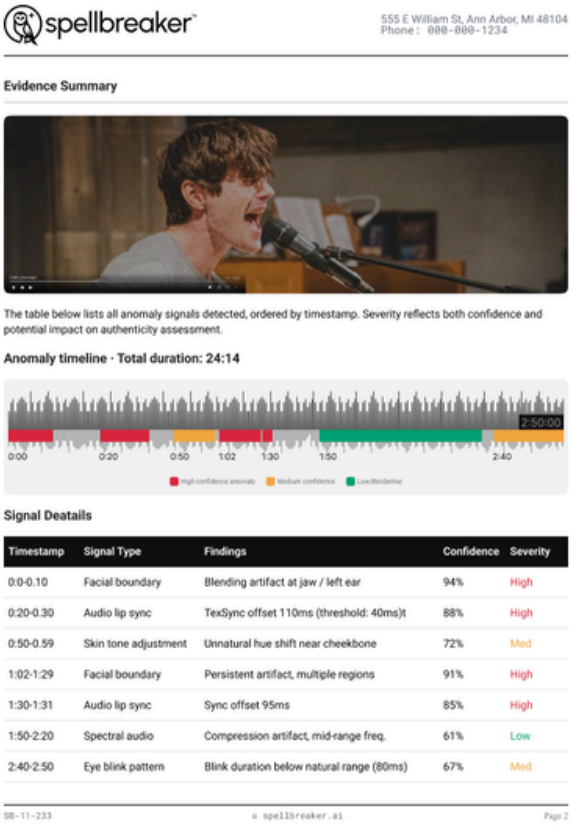
**What it means**

The combination of facial boundary artifacts, sync deviation, and audio-visual inconsistency strongly indicates that the face in this video has been replaced using a face-swap model, and the audio may have been separately synthesized or re-dubbed. The original speaker's identity cannot be confirmed from this recording.

**Limitations**

Our models are optimized for common face-swap and voice synthesis techniques. Novel generation methods introduced after our last model update (February 2026) may not be detected. Compression at source can reduce detection sensitivity. This report reflects the state of the provided file and does not account for re-encoding that may have occurred after manipulation

SB-11-233 spellbreaker.ai Page 1



spellbreaker

555 E William St, Ann Arbor, MI 48104
Phone: 000-000-1234

**Evidence Summary**

The table below lists all anomaly signals detected, ordered by timestamp. Severity reflects both confidence and potential impact on authenticity assessment.

**Anomaly timeline · Total duration: 24:14**

0:00 0:20 0:50 1:02 1:30 1:50 2:40 2:50:00

**Signal Deatails**

| Timestamp | Signal Type | Findings | Confidence | Severity |
|---|---|---|---|---|
| 0:0-0.10 | Facial boundary | Blending artifact at jaw / left ear | 94% | High |
| 0:20-0.30 | Audio lip sync | TexSync offset 110ms (threshold: 40ms)t | 88% | High |
| 0:50-0.59 | Skin tone adjustment | Unnatural hue shift near cheekbone | 72% | Med |
| 1:02-1:29 | Facial boundary | Persistent artifact, multiple regions | 91% | High |
| 1:30-1:31 | Audio lip sync | Sync offset 95ms | 85% | High |
| 1:50-2:20 | Spectral audio | Compression artifact, mid-range freq. | 61% | Low |
| 2:40-2:50 | Eye blink pattern | Blink duration below natural range (80ms) | 67% | Med |

SB-11-233 spellbreaker.ai Page 2

The reporting experience was redesigned to make findings easier to understand, more actionable, and better suited to different user needs. The earlier report felt overly technical, with information spread across sections that were difficult to interpret quickly, especially for non-expert users. The new structure leads with the verdict, confidence score, and a plain-language summary before moving into supporting evidence and deeper technical detail.

The final design includes both basic and detailed reports. The basic report supports quick understanding and decision-making, while the detailed report provides deeper forensic insight for expert review. Across both formats, the report is more structured, visually prioritized, and easier to scan, making it feel more transparent, credible, and aligned with how different stakeholders review forensic evidence.

# Realization Plan

Turning this prototype into a shipped product requires a phased approach. Each phase builds on the last, prioritizing the highest-impact changes first while laying the technical foundation for enterprise-scale deployment.

## Phase 1: 0 - 3 months

Core report redesign: Implement basic and detailed report formats. Deploy confidence legend and chain-of-custody metadata. Introduce a verdict-first information hierarchy. Replace legacy report structure across all user-facing outputs.

## Phase 2: 3 - 6 months

Dashboard & Upload overhaul: Unified analysis history table with thumbnail previews and role-based access. Rebuild upload flow with in-product compression, trimming, and constraint visibility. Case name field and report type selection at upload.

## Phase 3: 6 - 12 months

Enterprise Auth & API Hardening: Deploy SSO, MFA, magic link, and org-level permissions. Admin dashboard with cross-user file visibility. API-first architecture to support court management system and AAA arbitration platform integrations.

## Phase 4: 12 months and beyond

Scale, localization & expanded detection for other language output. On-premise deployment for government and enterprise clients. Liveness detection for eKYC use cases. Object-level manipulation detection for insurance fraud. Cryptographic provenance and audit trail integration.

# Conclusion

## Summary

Spellbreaker must evolve from a detection tool into a legal-grade credibility system, one that supports authentication, explainability, and procedural reasoning at every step. Trust in AI-generated evidence is no longer about visual realism. It is constructed through process, provenance, and defensibility. Our redesign lays the foundation, but long-term success depends on continued validation with practicing legal professionals.

Recommendation 01: Design every report as a legal document, assuming it will be cross-examined. Vague language, undefined terms, and unexplained scores are liabilities, not usability issues.

Recommendation 02: Use calm, institutional language throughout. Avoid overconfident claims. Conservative framing and visible limitations build more trust than certainty.

Recommendation 03: Surface limitations visibly. Transparency about what the system cannot determine is just as important as what it can. Inconclusive results must be communicated clearly with next-step guidance.

Recommendation 04: Invest in deterrence as a product narrative. If legal professionals know courts use Spellbreaker, fraudulent evidence submissions decrease before they happen, a powerful second-order benefit not yet communicated in the product.

Recommendation 05: Prioritize follow-up research with practicing attorneys and judges to validate design decisions with the primary target audience. Law students provided valuable feedback, but the real test is with courtroom users.