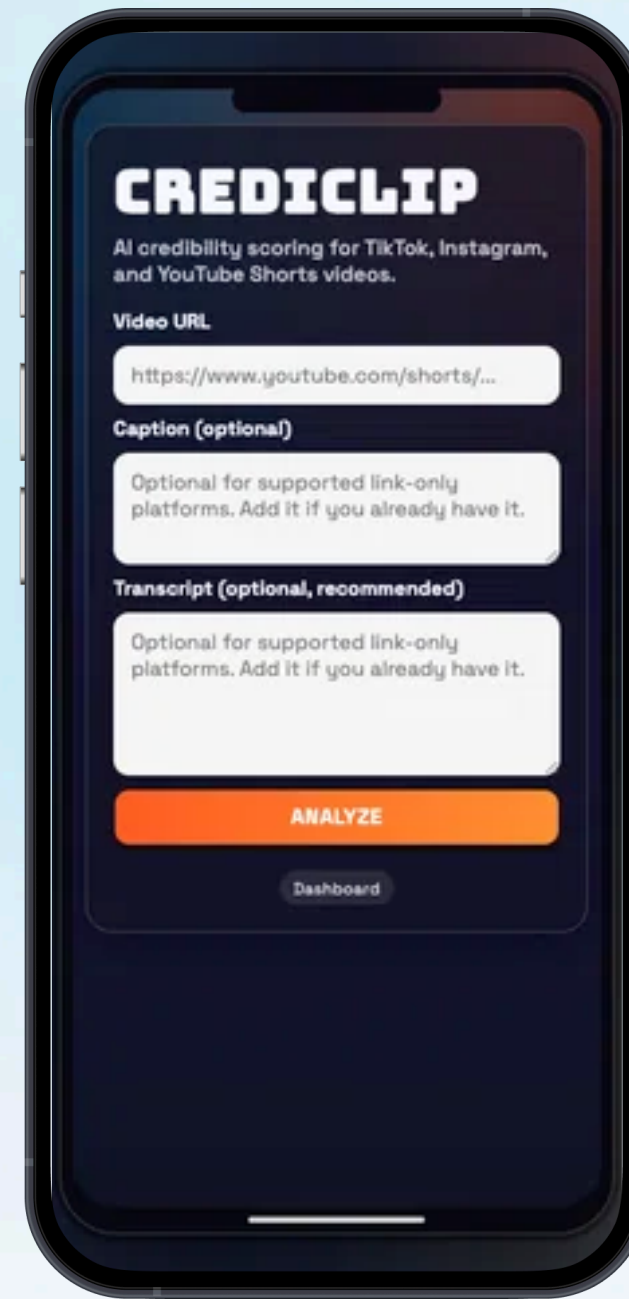


CrediClip

AI-powered credibility analysis for short-form video

Turns public video links into simple, explainable credibility signals.
Supports Instagram, TikTok, and YouTube Shorts.



CREDICLIP

AI credibility scoring for TikTok, Instagram, and YouTube Shorts videos.

Video URL

<https://www.youtube.com/shorts/...>

Caption (optional)

Optional for supported link-only platforms. Add it if you already have it.

Transcript (optional, recommended)

Optional for supported link-only platforms. Add it if you already have it.

ANALYZE

Dashboard





The problem with short-form video

01

Misinformation

False or distorted claims spread before anyone can verify the source or context.

02

Scam

Deceptive intent hidden behind polished content, fake offers, and manufactured urgency.

03

Manipulation

Emotional and psychological pressure used to bypass critical thinking and drive reactions.

04

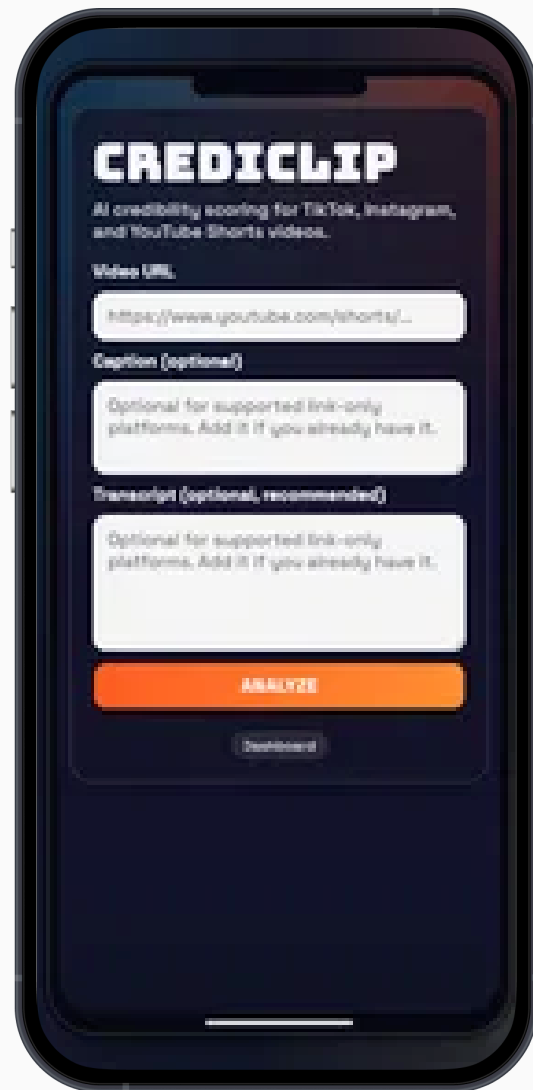
Uncertainty

Unclear origin and weak context leave viewers without the signals needed to trust what they see.

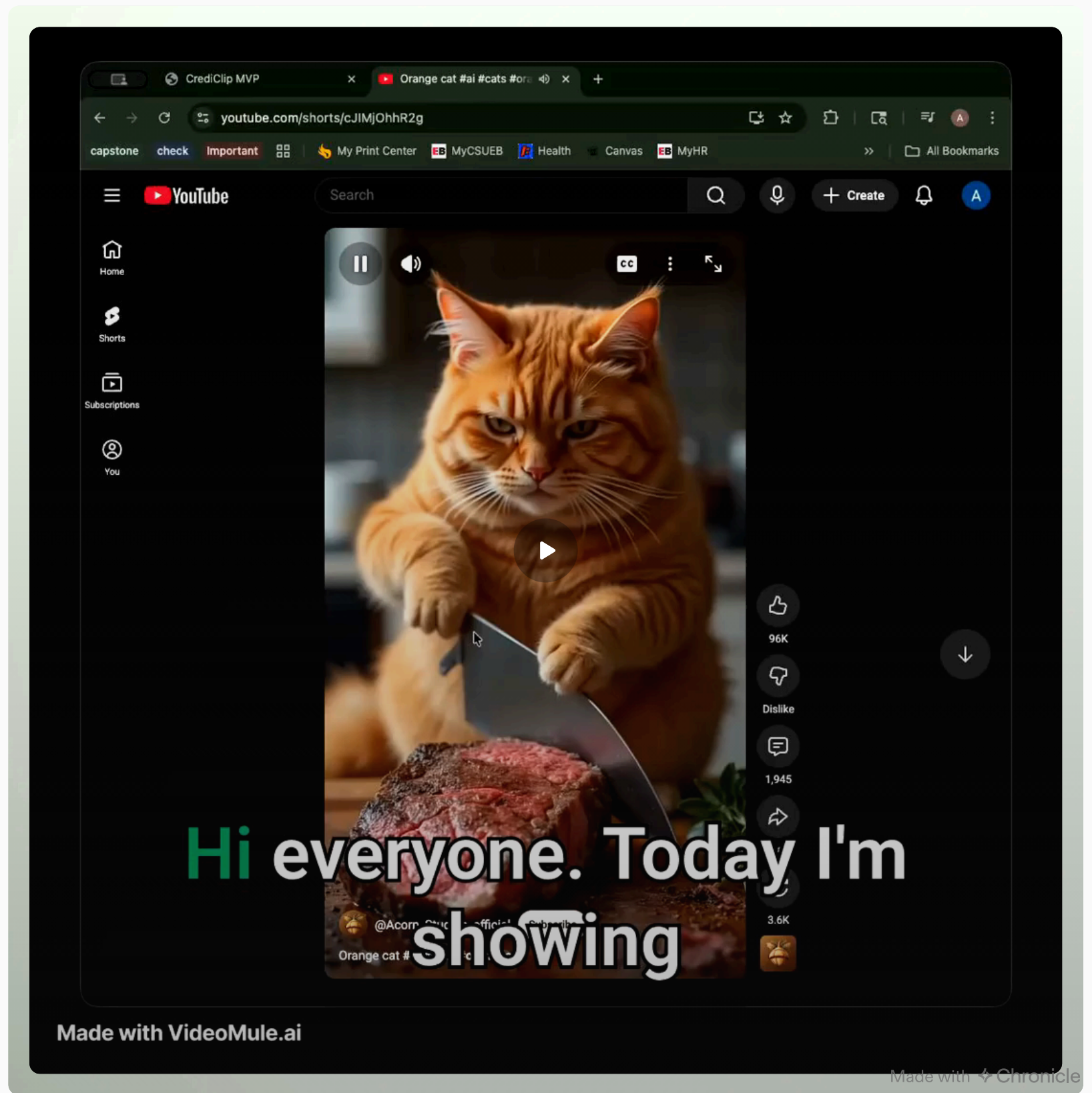
- | | | |
|---|------------------|---|
| 1 | Paste a link | User submits any short-form video URL to begin analysis. |
| 2 | Detect platform | System identifies whether the source is Instagram, TikTok, or YouTube Shorts. |
| 3 | Extract evidence | Workers pull metadata, captions, transcript, OCR text, and rendered content. |
| 4 | Score signals | Model evaluates risk across misinformation, scam, AI content, manipulation, and evidence quality. |
| 5 | Show result | User receives a clear, explainable credibility score — no technical knowledge needed. |

CrediClip

Live Demo



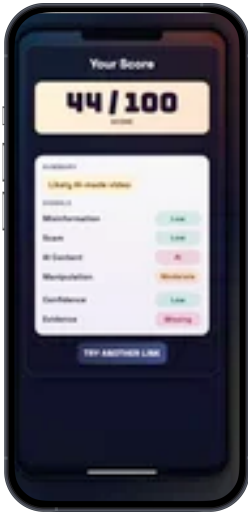
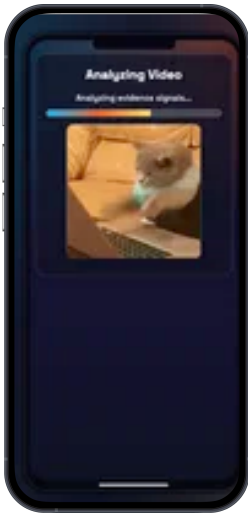
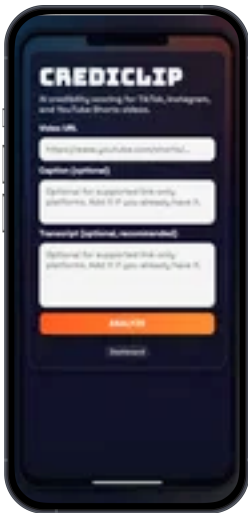
Try Now



What happens behind the scenes

CrediClip doesn't guess.

It pulls structured evidence from a video link, runs it through parallel workers, and feeds everything into a deterministic scoring engine.



 Frontend / API

Receives the video link and initiates the request

↓

 Queue

Manages jobs and prevents overload

↓

 Workers

Parallel media processing: metadata, OCR, transcript, captions

↓

 Evidence bundle

All extracted signals packaged into a single structured object

↓

 Scoring engine

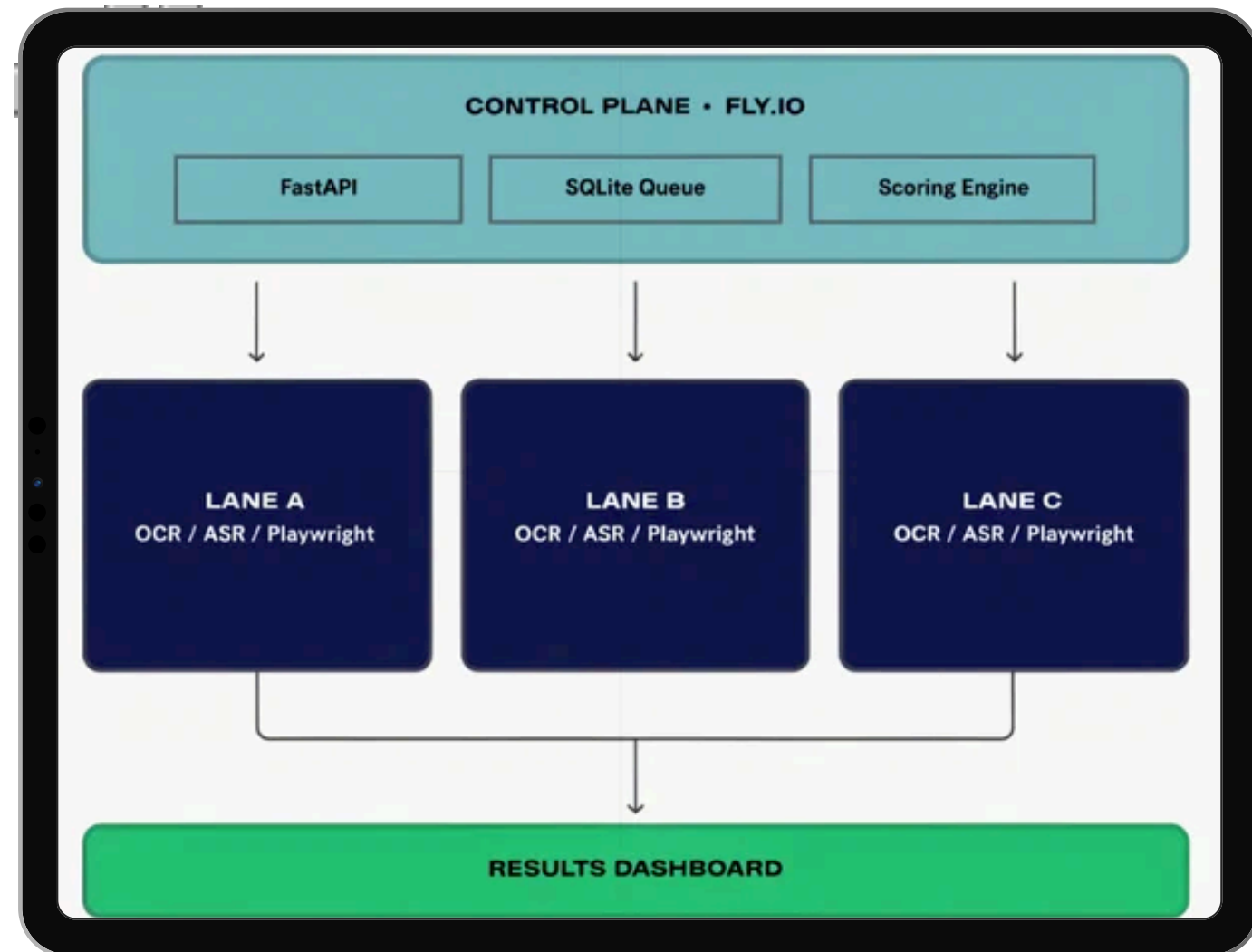
Combines all signals into a final explainable credibility score

System Architecture

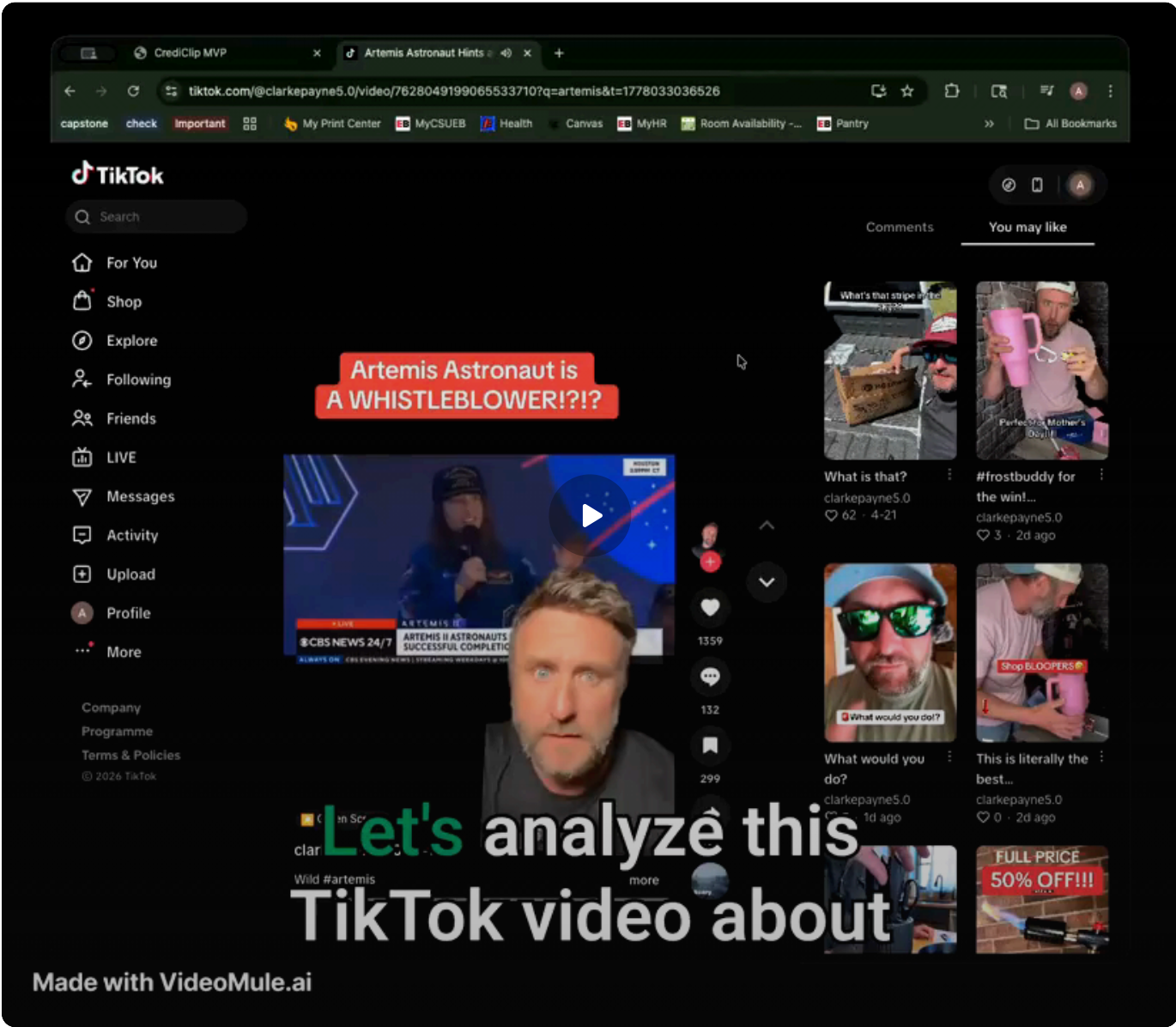
Control plane — Fly.io
FastAPI · SQLite queue · Scoring engine · Dashboard UI

Data plane — Oracle VMs
Lane A & B: Playwright, Tesseract
OCR, faster-whisper ASR, yt-dlp
Lane C: YouTube-only · ffmpeg ·
browser fallback · OCR · ASR

Flow
User → FastAPI → Queue → Workers →
Evidence extraction → Scoring →
Results



Video analysis



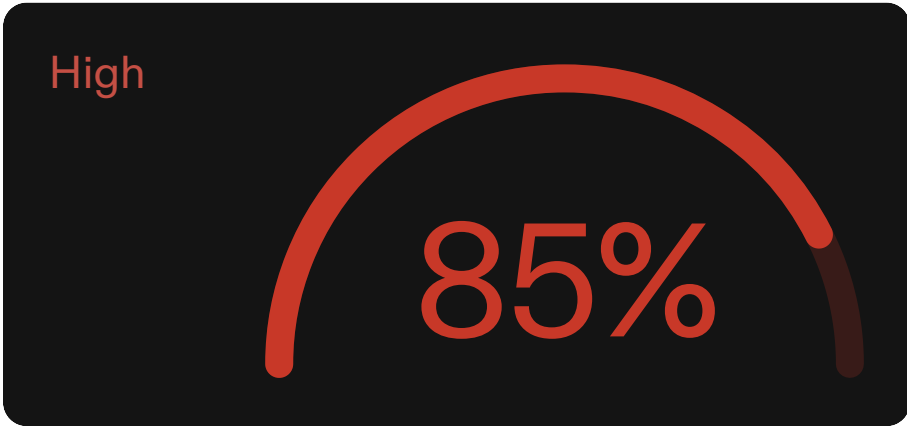
CrediClip result

Score: 36 / 100 —
Significant **Misinformation** signals detected.

Scene script for AI video generation:

- vertical TikTok explainer video about space or NASA Artemis
- confident narrator presenting a bold claim
- dramatic space footage, rocket launch visuals, moon imagery
- authoritative tone but questionable framing
- text overlays presenting claim as fact
- overall feel: persuasive but potentially misleading information clip

Misinformation Evidence is High.

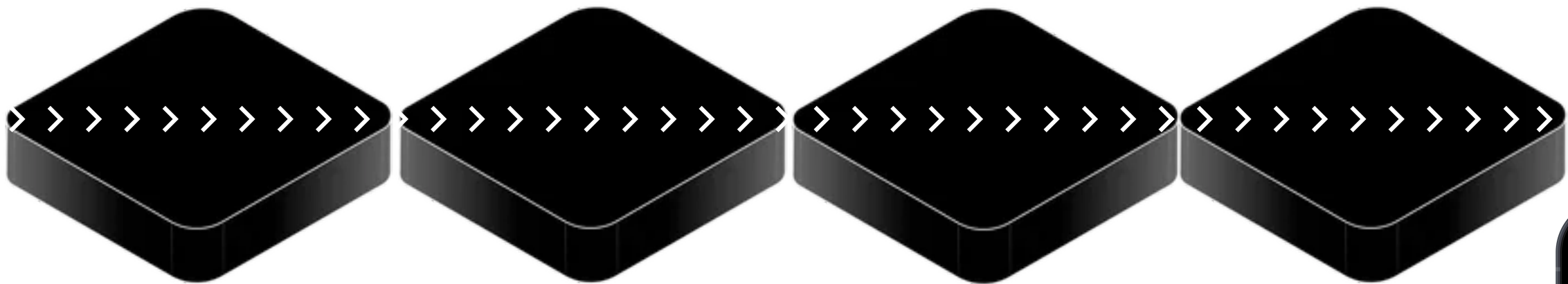


● Low risk ● Moderate risk ● High risk

Behind the Link



TikTok



Input + Data

- User pastes video link
- Platform data from YouTube, TikTok, Instagram

Uses live evidence from:

- metadata
- captions / transcript
- OCR / visuals
- browser-rendered content

Reference + Training

Uses benchmark/reference datasets:

- FEVER
- FaceForensics++
- DFDC
- ASVspoof
- AI or Not

Uses labeled training/reference examples

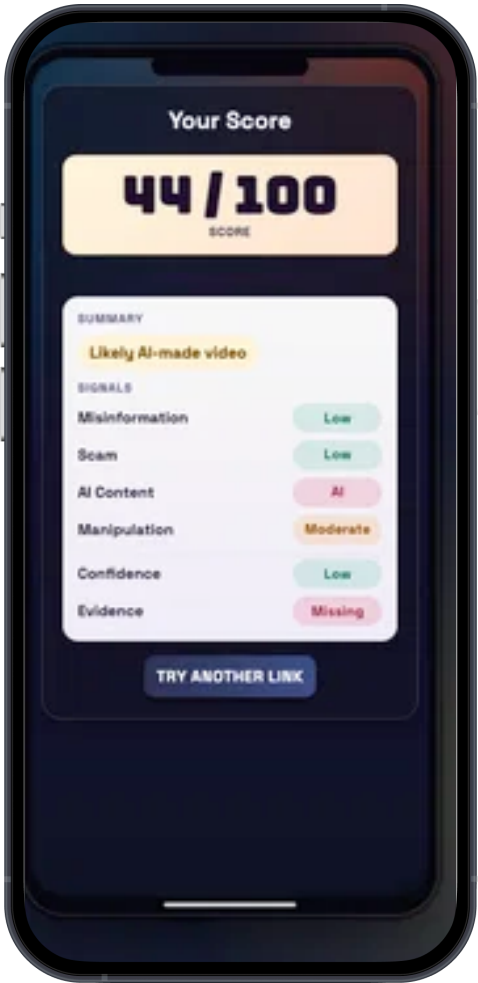
- Supports calibration for AI vs human and risk signals
- Labeled training data

Analysis Models

- LLM supports claim reasoning when evidence is strong enough
- Platform pipelines combine extracted evidence into one analysis bundle
- Heuristic model checks scam, misinformation, and manipulation cues

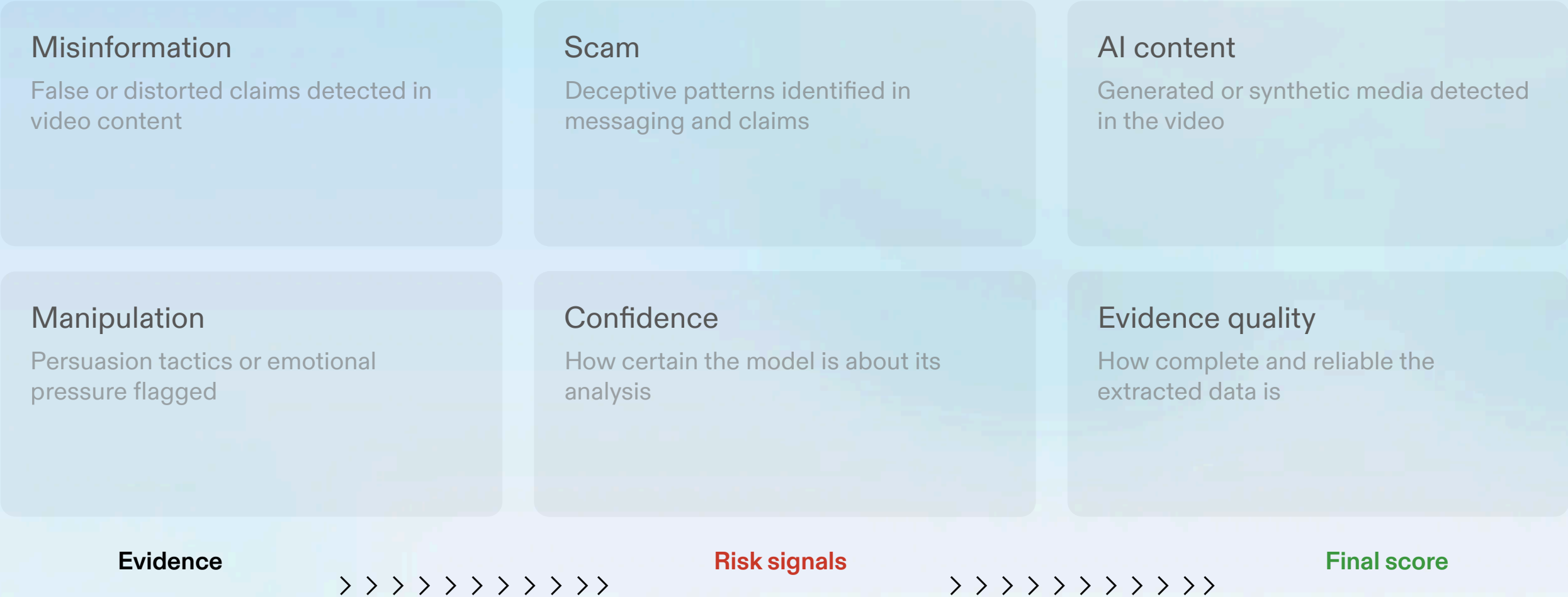
Scoring + Output

- Misinformation
- Scam
- AI / Manipulation
- Confidence
- **Final score**



How the score works

Six signals. One deterministic result.



High-score video

Score: 88.81 / 100 —

Strong, complete evidence

Clear signals and minimal risk cues raise confidence.

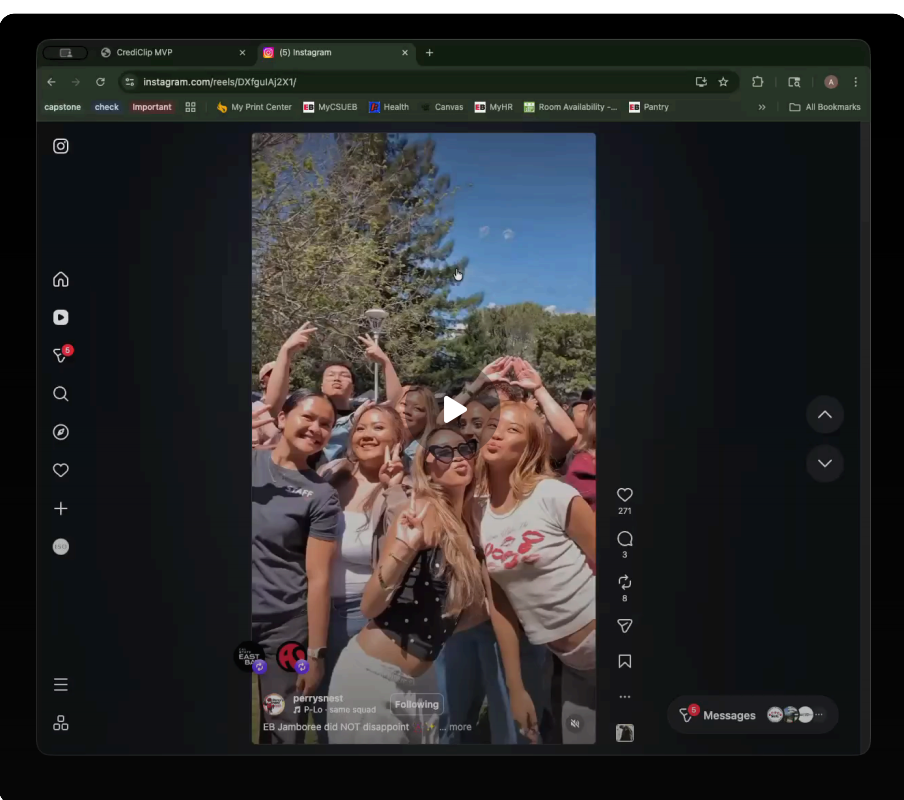
Result: trustworthy

1 . Comprehensive evidence found

2 . Low risk signals detected

3 . High confidence in verdict

4 . Platform yields rich data



Low-score video

Score: 34 / 100 —

Weak or missing evidence

AI, scam, or manipulation cues increase uncertainty.

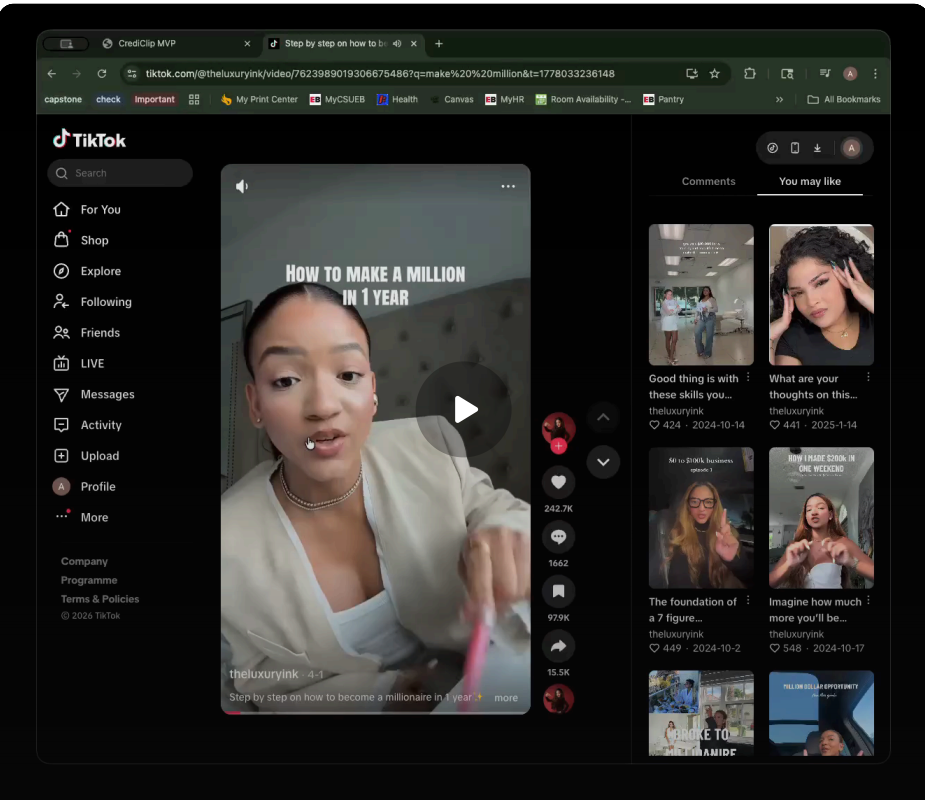
Result: flagged or uncertain

1 . Sparse or unavailable evidence

2 . Stronger risk signals present

3 . More uncertainty in analysis

4 . Platform limits data extraction



Signal patterns by category

How different video types produce different credibility signals across CrediClip's analysis dimensions.

Category	Score direction	Strongest signal	Evidence quality	Result
AI-generated	▼ Low	AI content flag	Weak	🚩 Flagged
Human / authentic	▲ High	Low risk — all signals	Strong	✓ Safe
Scam	▼▼ Very low	Scam indicator	Thin / misleading	🚩 Flagged
Misinformation	▼ Low-medium	Misinfo flag	Partial	⚠ Caution
Green flag / high score	▲ High	All signals low risk	Rich	✓ Credible

←
High Score Video Demo 1

←
Low Score Video Demo 2

Why CrediClip matters

Link in. Evidence out. Clearer trust signal.



Business prospect

- Moderation triage: surfaces suspect video with evidence-backed reasoning
- Trust and safety API: credibility layer for platform workflows
- Fact-checking support: faster screening of high-volume video
- Media literacy: transparent credibility framing for end users