HYPOTENUSE ANALYTICS

One Platform Three Signals
One Truth

Predict. Protect. Verify.

ZSure — Reality Trust Center Technical White Paper

Contents

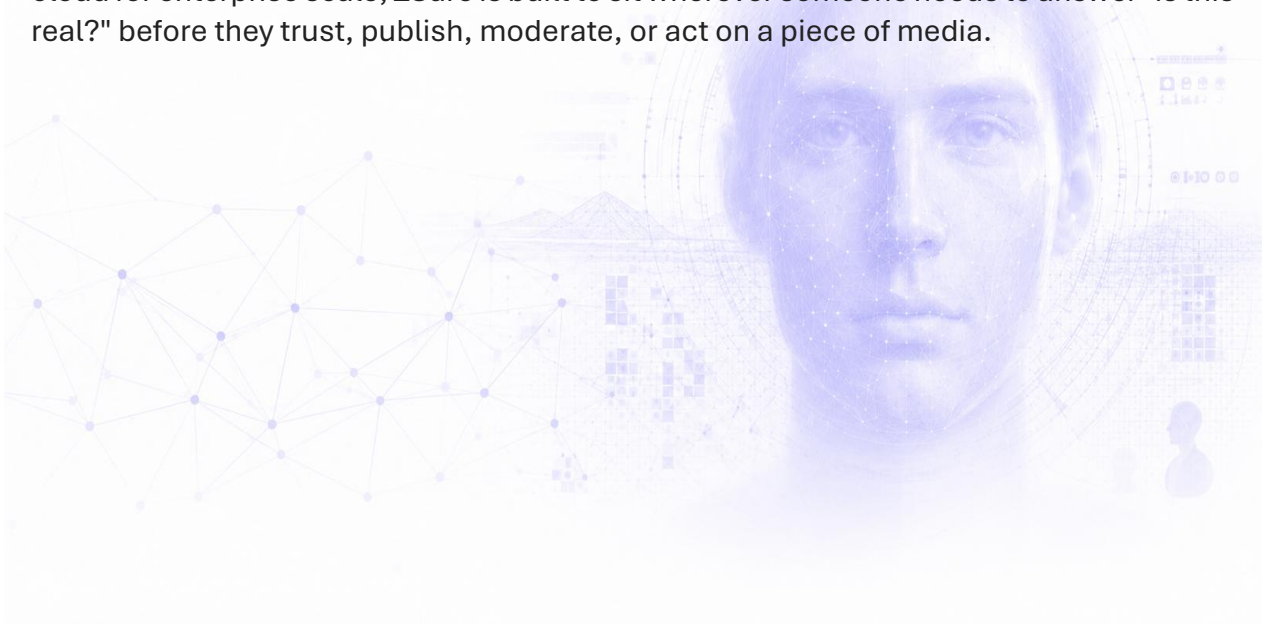
Introduction

Synthetic images and video are now good enough to fool the naked eye — and cheap enough that anyone can generate them. .ZSure, the detection engine inside the Reality Trust Center, takes in an image or video from anywhere — a chat app, a social post, a news article, a raw upload — and returns an AI-generation probability score backed by visual evidence in seconds, not a bare **real/fake** label.

Instead of betting on one AI model, ZSure runs **three independently-trained detection models** at once and fuses their opinions into a single score — closer to a panel of forensic experts than a single black-box classifier. When a new AI generator launches tomorrow that none of the three has seen before, the odds all three are fooled together are far lower than the odds any one of them is.

This matters because the problem isn't narrow. ZSure isn't only a deepfake detector in the sense of catching manipulated video of a real person — at its core, it's an **AI-or-not detector**: it determines whether any image or video was produced by a generative model at all, regardless of subject, from a synthetic product photo to an AI-generated avatar to a fully AI-written document. Deepfake detection — specifically catching when a real person's likeness has been synthesized or manipulated — is one important, high-stakes case that sits inside that broader detection capability, and both draw on the same three-model ensemble.

ZSure also isn't a standalone tool sitting in isolation. As part of the Reality Trust Center, it can cross-check a flagged video or document against real camera footage or sensor data from the same site or incident — something a company selling deepfake detection alone cannot do. Combined with deployment options that span a one-click browser badge, a REST API for product teams, direct upload for forensic casework, and serverless cloud for enterprise scale, ZSure is built to sit wherever someone needs to answer "is this real?" before they trust, publish, moderate, or act on a piece of media.



How It Works — The Three-Model Ensemble

Model	Role	Best At
Structure Detection Model	Checks whether the image's spatial "story" holds together the way a real photo's would	Catching never-seen-before / novel generators
Texture Detection Model	Looks closely at pixel-level noise and texture patterns	GAN-generated faces, classic face-swap fakes
Specialist Routing Model	Routes each image to internal specialist sub-experts, with one always checking for generic AI artifacts	Diffusion-model output (modern image/video generators)

Document Forgery Detection

Documents are a distinct forensic problem from video and voice, not a variant of either — a spliced Aadhaar or PAN card, an AI-generated bank statement, or a fabricated inspection report is now something a generative tool can produce wholesale rather than edit from a real one. ZSure's document pipeline analyzes layout and font consistency, metadata integrity, and splice boundaries invisible to the naked eye, evaluated against the DocTamper, DocForgery v2, SynDoc, and GhostWriter benchmark datasets.

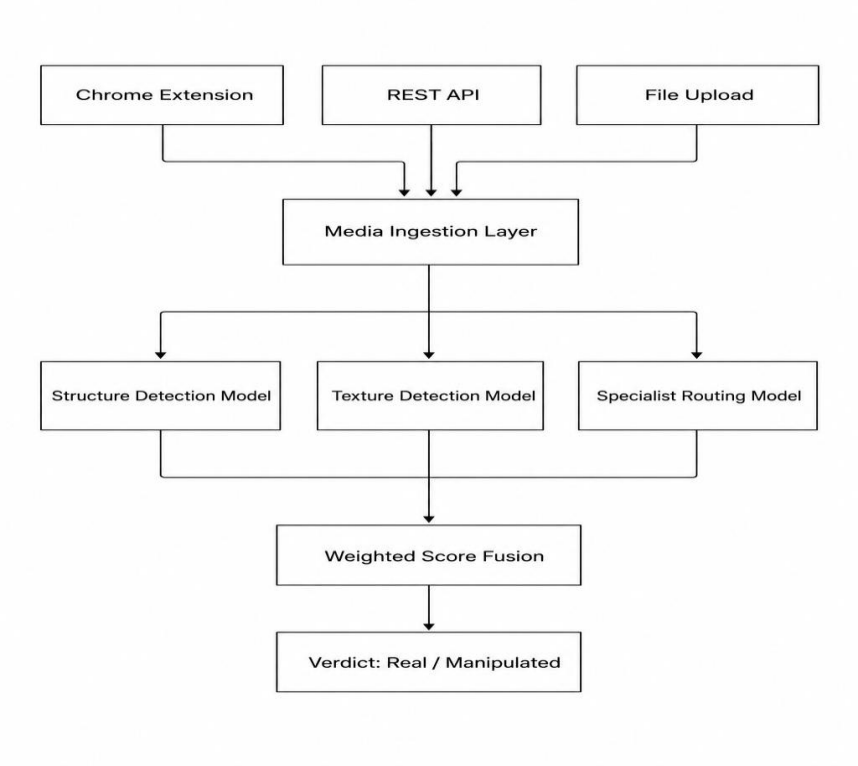
Voice Clone & Synthesis Detection

Cloned voices are now used in real time during urgent call-center transfer requests, so upload-and-analyze checks that run after a call has ended do not help. ZSure's voice pipeline uses a speech-embedding front end paired with a lightweight classifier to detect cloning and synthesis in near-real-time audio streams.

Why the Ensemble Works

The models don't just get averaged blindly — the system is weighted toward whichever model generalizes best against unfamiliar generators, with the other two acting as corroborating specialists. If all three models agree, the verdict is high-confidence. If they disagree, the result is flagged as ambiguous rather than forced into a guess — the system would rather say "uncertain" than give a false sense of certainty.

> **Video analysis:** the system samples frames adaptively across a clip, scores each independently, and combines them in a way that filters out noisy or blank frames — so a few bad frames can't flip the verdict.



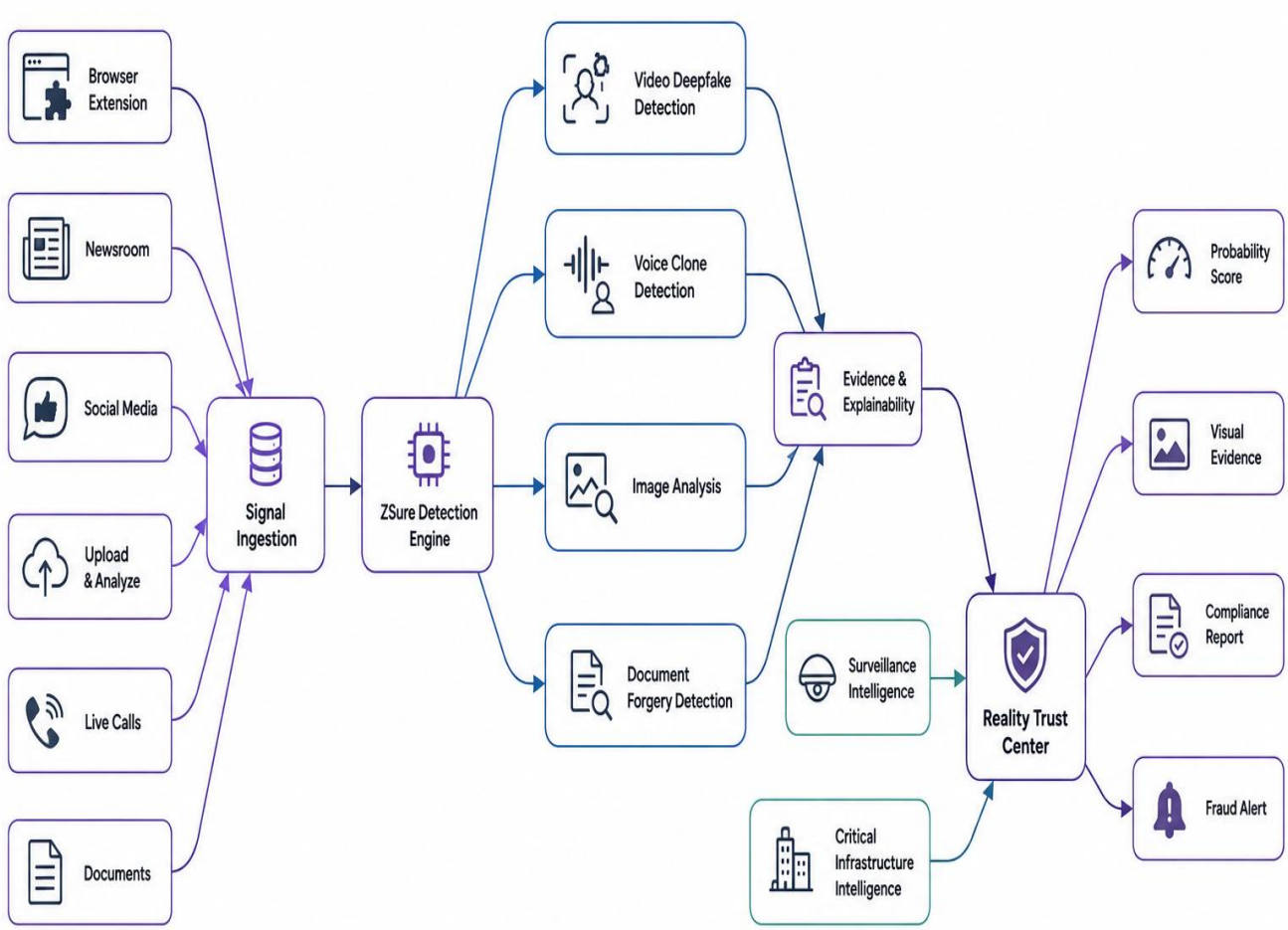
Live Video-Call Protection

Deepfake video is increasingly used to impersonate executives on live calls to authorize wire transfers or bypass verification — by the time an uploaded file is reviewed, the fraud has already happened. ZSure's live-call protection is designed to deliver sub-150ms detection latency with zero added call latency, using a dual-stream consistency model to catch face-swap-over-real-audio and voice-clone-over-real-video attacks that a single-modality detector misses.

Generator Attribution

Once content is flagged as likely synthetic, ZSure's attribution layer identifies the probable source generator — Midjourney, OpenAI's image models, Google Gemini, Stable Diffusion, and other major platforms — via generator-specific frequency fingerprints, rather than stopping at a bare synthetic/real call. Where the fingerprint doesn't match a known generator, ZSure reports "unknown/novel generator" rather than forcing a guess, consistent with the same honesty-over-false-confidence principle used when the three-model ensemble disagrees.

Architecture



This architecture shows how the Reality Trust Center processes inputs from a wide range of sources — browser extensions, newsrooms, social media, direct uploads, live calls, and documents — through a unified Signal Ingestion layer into the ZSure Detection Engine. Based on the content type, the engine performs video deepfake detection, voice clone detection, image analysis, or document forgery detection, routing each input to the right combination of the three detection models rather than treating every media type the same way.

Beyond the detection engine itself, the platform correlates its findings with Surveillance Intelligence and Critical Infrastructure Intelligence, so a flagged video or document isn't assessed in isolation — it can be cross-checked against real camera footage or sensor data from the same site or incident. The result is a single outcome that combines a probability score, visual evidence, and a fraud or compliance alert, giving investigators and moderation teams something they can act on directly rather than a bare number they have to interpret on their own.

Proven Performance — The Numbers

Results below are from controlled benchmark evaluations, not live production traffic — measured across 2,660 images spanning 5 independent public and internal benchmark datasets, each testing a different real-world condition (cross-generator generalization, real-world image quality, video frames, compression, and classic GAN faces).

Metric	Value
Accuracy	90.9%
Precision	90.3%
Recall	90.2%
F1 Score	90.3%
AUC-ROC	0.942

Per-Dataset Breakdown

Dataset	N	Accuracy	AUC-ROC	What It Tests
Cross-Generator Benchmark	2,000	**96.3%**	0.989	Multi-generator forensic benchmark
Hard Real-World Set	215	**92.6%**	—	Complex lighting, filters, screenshots
AI Video Frames	20	90.0%	—	Modern video-generator output
Compressed Social Media	400	63.2%	0.653	Heavily re-encoded/compressed content (active-improvement area)
Classic GAN Faces	25	92.0%	—	Synthetic-face-style benchmark

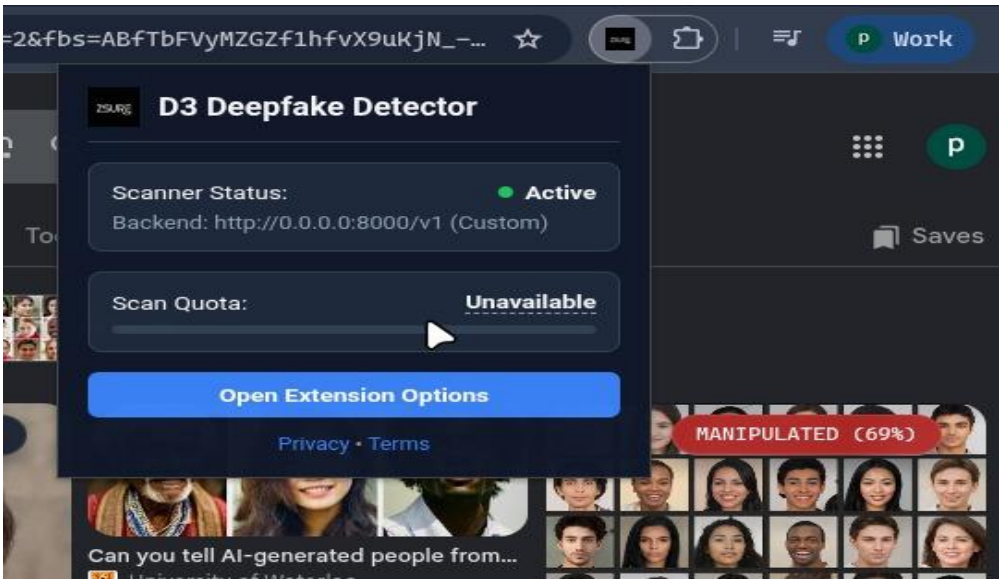
Evaluation prioritizes zero-shot performance — accuracy against generators the model has never seen during training — over in-domain accuracy on familiar data, since the real test of a detector is how it performs against next month's new model release. A large gap between in-domain and out-of-domain accuracy is the signal that a model has memorized its training set rather than learned to generalize; ZSure's cross-generator benchmark score is reported specifically to surface that gap rather than hide it.

Robustness is tested beyond compression alone: documents are validated against scan-print-scan cycles and low-DPI artifacting (since real fraud documents are photocopied or scanned, not passed around as clean PDFs), and detection is stress-tested against adversarial perturbations engineered to fool the model. Contrastive learning pushes real and fake samples further apart in embedding space regardless of which generator produced the fake, improving generalization to generators not seen during training.

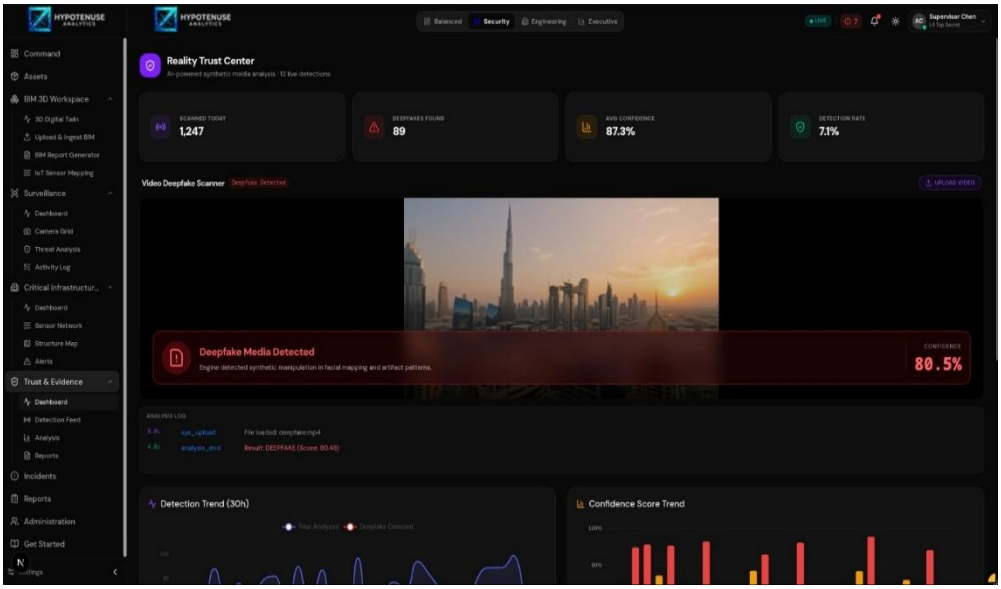
Reality Trust Analysis — detection type distribution across manipulation categories (left) and per-model accuracy (precision, recall, F1) by detection type (right), based on ZSure's internal benchmark evaluation.

How Customers & Users Access It

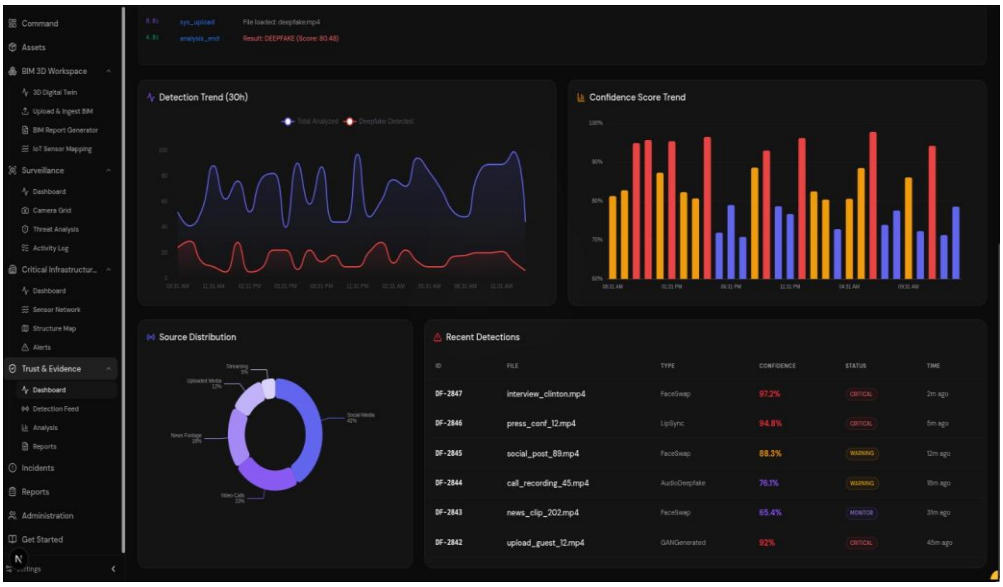
Touchpoint	For	How
Chrome Extension	End-users, journalists, fact-checkers	In-browser scan badge on images/video — green/red verdict with confidence %
REST API	Product teams, moderation pipelines	Structured JSON verdict, confidence score, per-model breakdown
Direct File Upload	Forensic investigators, security analysts	Full multi-frame video scanning with adaptive sampling
Cloud & Batch Engine	High-volume platforms	Serverless, auto-scaling GPU deployment for enterprise scale



Reality Trust Analysis — detection type distribution across manipulation categories (left) and per-model accuracy (precision, recall, F1) by detection type (right), based on ZSure's internal benchmark evaluation.



Reality Trust Center dashboard — live video deepfake scan showing a detected result with confidence score, alongside real-time platform stats (media scanned, deepfakes found, average confidence, detection rate).



Reality Trust Center dashboard (continued) — detection trend and confidence score trend over time, source distribution of scanned media, and a live log of recent detections with type, confidence, and status.

Competitive Positioning

Capability	Legacy Detectors	ZSure (Reality Trust Center)
Cross-Generator Accuracy	68–78% (fails on unseen generators)	**90.9%, 96.3%** on primary benchmark
Robustness to Compression	Significant degradation	Resilient (92.6% on hard real-world set)
Architecture	Single model	**3-model weighted ensemble**
User Experience	CLI tools / upload portals	1-click in-browser overlay + REST API
Deployment	Static server requirement	Local GPU, serverless cloud, or hybrid
Cross-referencing	None — standalone tool	Correlates with live camera/sensor data via Reality Trust Center
Source/Generator Attribution	None — flat real/fake only	Identifies likely source generator (Midjourney, OpenAI, Gemini, Stable Diffusion, etc.), not just a verdict

Key USPs & Highlights

Three-Model Ensemble, Not a Single Black Box — three independently-trained models fused into one weighted verdict; when one is fooled, the others catch it.

Strong Cross-Generator Accuracy — 96.3% (AUC 0.989) on the primary cross-generator benchmark, versus 68–78% for legacy single-model detectors.

Built for Real-World, Indian Media Conditions — 92.6% accuracy on hard real-world photos (tricky lighting, heavy filters, screenshots), tuned for typical Indian mobile/messaging compression.

Deploy Anywhere Media Shows Up — Chrome badge, REST API, direct upload, and serverless enterprise cloud, all remote-config driven.

Part of a Bigger Trust Fabric — Plugs into the Reality Trust Center, so a flagged video or document can be cross-checked against real camera footage or sensor data from the same site/incident — something a standalone deepfake detector can't do.

Privacy-Preserving Federated Learning — Each deployment learns from the media it checks; only learned patterns are shared across sites, never raw media. Poisoning is treated as a day-one threat, with updates anchored against confirmed cases and outliers rejected before reaching the shared model.

Tri-Layer Explainability Engine — Highlights the exact pixels behind a fake-probability score, checks whether the model relies on learned concepts rather than surface artifacts, and uses frequency analysis to catch fingerprints invisible to the eye.

Provenance-Aware — ZSure checks for open-standard provenance metadata where present. Its absence is treated as a signal worth noting, not proof of manipulation — most content in circulation today carries no provenance metadata at all.

Identifies the Source Generator — Beyond flagging that content is AI-generated, ZSure identifies which platform likely produced it — including major generators such as Midjourney, OpenAI, Google Gemini, and Stable Diffusion — giving investigators a lead on the source, not just a verdict.

Conclusion

ZSure combines three complementary detection models into a single weighted ensemble that holds up across a wide range of AI generators, including ones it has never seen before — closer to a panel of forensic experts than a single black-box classifier. Backed by evaluation across multiple independent benchmark datasets, it performs strongly on cross-generator detection while remaining honest about where it still has room to improve, such as heavily compressed social media content.

Deployable from a one-click browser badge to a REST API, direct forensic upload, and enterprise cloud infrastructure, ZSure is built to sit wherever someone needs to answer "is this real? " — and as part of the Reality Trust Center, it goes a step further, cross-checking flagged media against real camera footage or sensor data from the same site or incident, something a standalone deepfak e-detection tool cannot do.

CONTACT & CONNECT

INDIA OFFICE

Innov8 Graphix 2, sector 62, Ground Floor, Graphix Tower 2, A-13, Sector 62, Noida , Uttar Pradesh, India – 201301

info@hypotenuseanalytics.com

+91 63884 33626

FOLLOW HYPOTENUSE ANALYTICS

FOLLOW ZSURE®