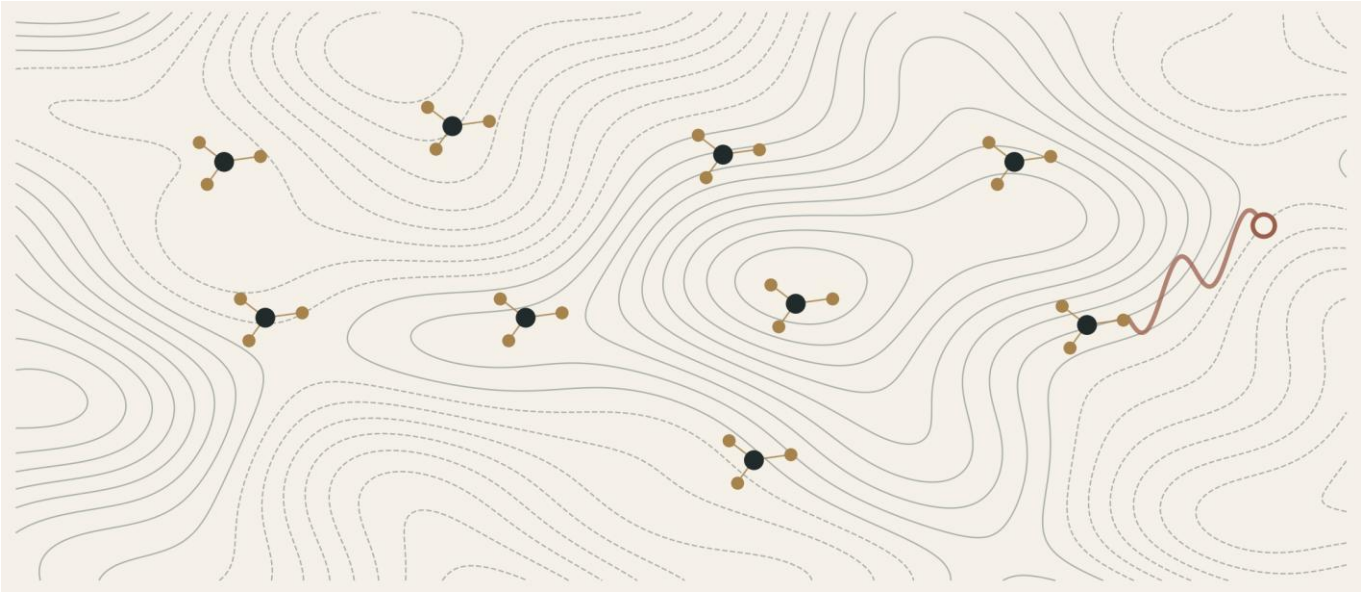
GROUNDTRUTH AI

Proof of Evidence

What has been demonstrated — and what the next investor-backed validation must prove



The transition: from a founder-built prototype to a reproducible regional intelligence system.

27	9	3	PASS
REGIONAL MODELS	GEOGRAPHIC REGIONS	INDEPENDENT RUNS / REGION	LEAKAGE-CONTROL CHECKS

Prepared by Ali Faraj, Founder | July 2026

Non-confidential: this brief excludes source code, exact sensitive coordinates, raw training data, model files, credentials, and proprietary feature recipes.

The proof in one minute

A clear reading of what changed, what is credible today, and what still needs to be tested independently.

Why I am sharing this now

GroundTruth AI completed a regional-expert training stage. The project is no longer represented by a single model result. It now has a repeatable architecture, a documented performance uplift, preserved model identities, and a practical route to external validation.

— Ali Riyad Faraj

The one-sentence conclusion

Internal technical risk has been reduced significantly. The next risk to test is whether the result holds on unseen data, with independent review and better imagery.

27	0.773–0.803	0.864–0.907	PASS
MODEL ARTIFACTS SAVED	REGIONAL AUC RANGE	REGIONAL RECALL RANGE	INTERNAL CONTROL STATUS

What has been shown

- The regional training pathway runs across nine regions.
- Three independent runs per region reduce reliance on one lucky outcome.
- Regional models outperform the reported global baseline ranges.

What this does not prove

- It is not yet a blind external test.

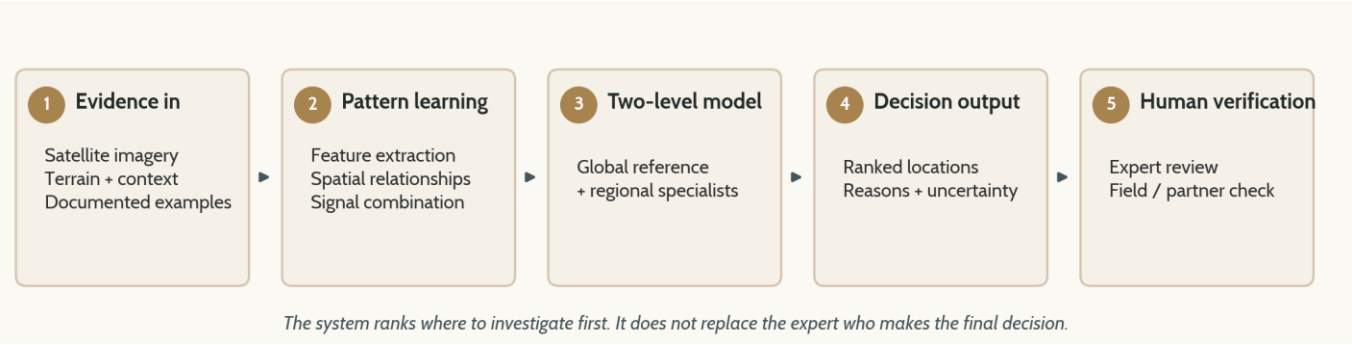
What the next step must prove

- Performance on partner-held unseen outcomes.
- Search-area reduction without losing important candidates.
- Clear value for expert and field decisions.

Investor reading: the project has moved beyond "interesting patent prototype," but it has not yet crossed into independent live proof. That is exactly what the next pilot should be designed to establish.

What GroundTruth AI actually does

The system is a prioritization layer: it narrows where experts should investigate first and explains why.



The system ranks where to investigate first. It does not replace the expert who makes the final decision.

The problem

- Large areas are expensive to inspect manually.
- Relevant patterns can be weak, scattered, and different by region.
- Field teams need a defensible shortlist before committing time and money.

The product

- Combines satellite, terrain, historical, and contextual evidence.
- Ranks locations rather than issuing an unsupported yes/no claim.
- Produces reasons, uncertainty notes, and an audit trail for review.

The human role

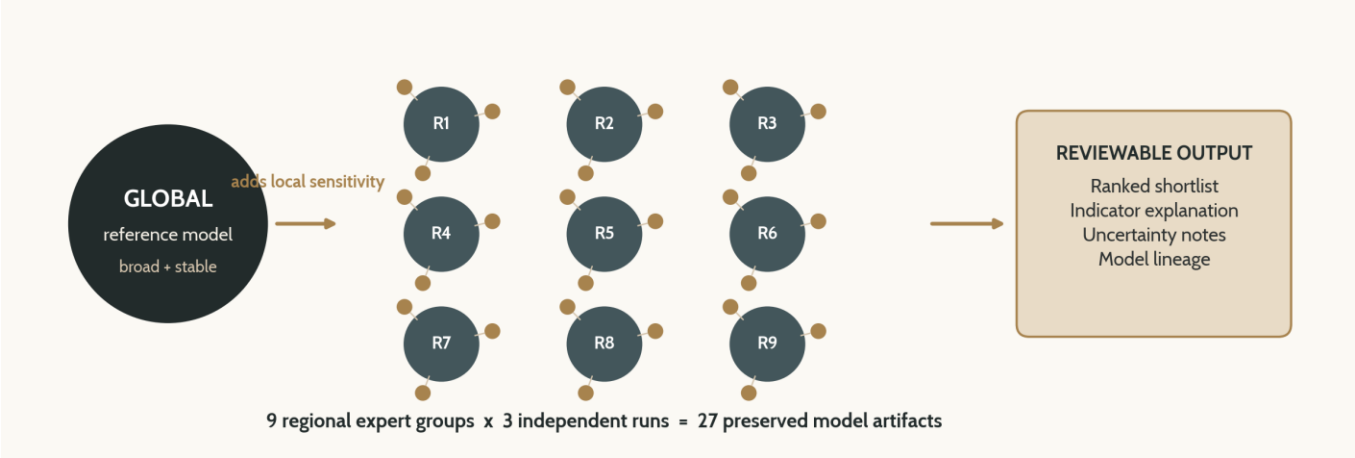
- Domain experts decide what is plausible.
- Partners control field confirmation and operational decisions.
- The model supports judgment; it does not replace it.

A useful analogy: the global model behaves like a general practitioner who keeps a consistent reference. The regional models act like local specialists who are more sensitive to patterns that change with terrain, surface conditions, and geography.

How the current evidence was built

A staged progression from data discipline to a repeatable regional-expert architecture.

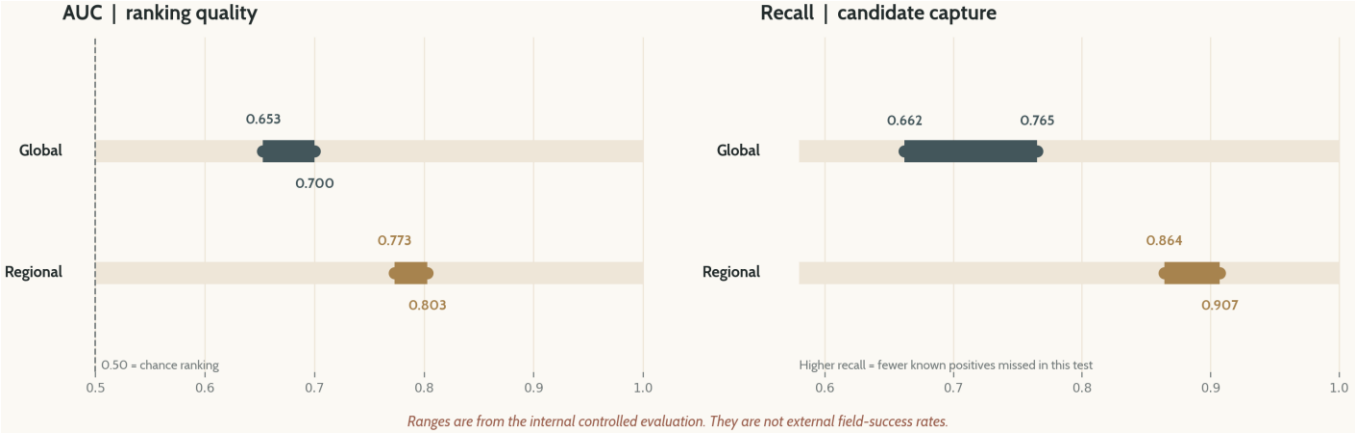
Stage	What was done	Why it matters
1. Data foundation	Documented positives, controlled negatives, provenance review, exclusions, and quality checks.	A defensible base before trusting any metric.
2. Controlled signal package	62 documented positives and 488 controlled negatives were used in an earlier signal experiment.	Evidence that the pipeline could learn more than random noise under controlled conditions.
3. Global reference model	A broad baseline was retained with AUC 0.653–0.700 and Recall 0.662–0.765.	A stable comparison point and governance layer.
4. Regional expert layer	Nine regions were each trained three times, producing 27 regional model artifacts.	Local sensitivity without abandoning the global reference.
5. Reproducibility record	Model fingerprints, result summaries, quota records, and control outcomes were preserved.	A reviewable trail for technical diligence.



Why the global model was not replaced: the regional models are an additional high-recall review layer. Keeping the global model preserves a stable reference for cross-region comparison, governance, and future external testing.

What the numbers actually mean

The most important improvement is not “a bigger number.” It is better ranking and fewer missed known positives in the controlled evaluation.



AUC = ranking quality

AUC asks how often the model ranks a known positive above a control example. A value of 0.50 is chance ranking; 1.00 is perfect ranking under that specific test.

Recall = candidate capture

Recall asks how many known positives were flagged. A regional recall range of 0.864–0.907 means roughly 86–91 out of every 100 known positives were captured in this controlled evaluation.

Global reference

- AUC: 0.653–0.700
- Recall: 0.662–0.765
- Broad and consistent across regions.

Regional expert layer

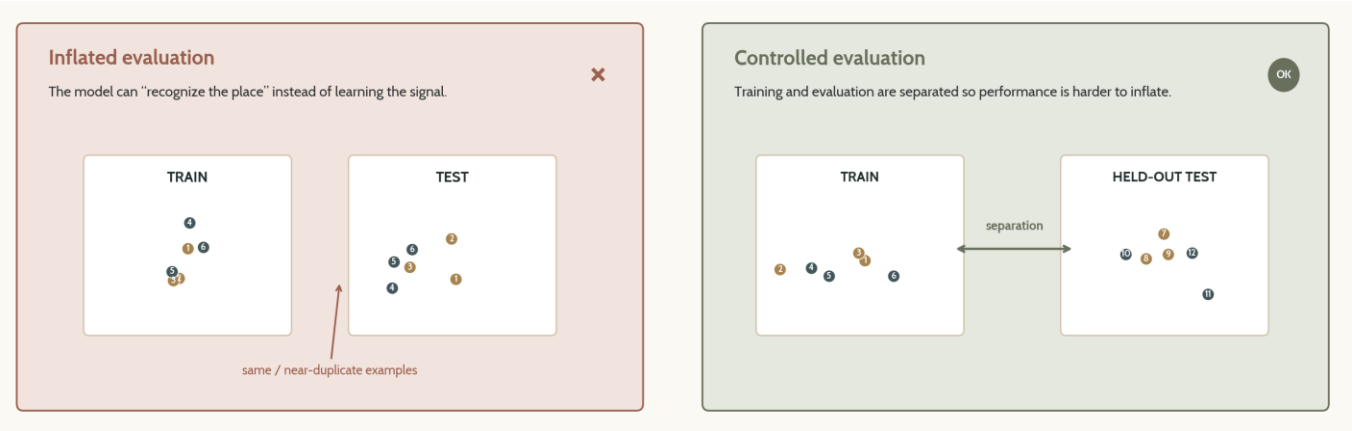
- AUC: 0.773–0.803
- Recall: 0.864–0.907
- Designed to catch more local candidates.

Correct interpretation

- Clear uplift over the reported baseline.
- F2 also improved, supporting the recall-oriented design.
- Still requires external blind validation.

Why leakage controls matter

A model can look impressive for the wrong reason if training and evaluation are not properly separated.



In plain language, "leakage" means the model is given an unfair hint. The hint may be a duplicated example, a nearby version of the same place, or information that indirectly reveals where the example came from. A leakage-prone test can reward memory instead of genuine pattern learning.

Control	What it means	Investor value
Leakage-prevention checks	PASS in the internal training record	Reduces the risk of an artificially inflated evaluation.
Three seeds per region	The same regional experiment was rerun from different random starting points	Makes the result less dependent on one lucky run.
Variable quotas	The planned sample allocation was applied by region, retaining all samples in smaller regions	Uses scarce regional evidence consistently.
27 model fingerprints	A unique digital checksum was saved for every model artifact	Allows reviewers to confirm which exact model produced a result.

Why this matters: the result is easier to review, reproduce, and challenge. It is not the same as an independent audit, but it is a stronger foundation for one.

What is proven — and what is not

Credibility comes from separating demonstrated facts from the claims that still need external validation.

Demonstrated now

- A reproducible nine-region training path.
- Twenty-seven preserved regional model artifacts.
- Reported uplift over the global baseline ranges.
- Leakage-control checks recorded as passed.
- A non-confidential evidence trail suitable for investor review.

Not claimed yet

- Guaranteed discovery in a new geography.
- Independent blind generalization.
- Commercial-imagery performance.

Next proof required

- Partner-held unseen outcomes.
- A sealed test area and locked protocol.
- Outputs produced before truth is revealed.
- Independent scoring and expert review.
- Field or documentary verification where appropriate.

The strongest defensible statement today

GroundTruth AI has demonstrated a governed regional-expert training pathway that improves candidate ranking and capture over its global reference model in the recorded internal evaluation. The next milestone is independent proof, not another founder-only test.

A useful investor distinction: proof of concept shows that a technical pathway works under controlled conditions. Proof of market requires independent performance, operational value, and a partner willing to use the output.

This boundary is deliberate. It protects the credibility of the project and makes the proposed next step measurable rather than promotional.

The practical ceiling of individual development

The project has reached the point where the next measurable gains depend on institutional inputs, not simply more founder hours.

The current milestone

This brief represents the highest technically defensible milestone achievable at the current founder-led, self-funded stage using publicly accessible imagery and independently available computational resources.

Built independently

- Data collection and provenance discipline
- Training and evaluation pipeline
- Global and regional model experiments
- Metric summaries and preserved model fingerprints
- Investor-readable non-confidential evidence pack

→ **NEXT STAGE**

Requires an institutional layer

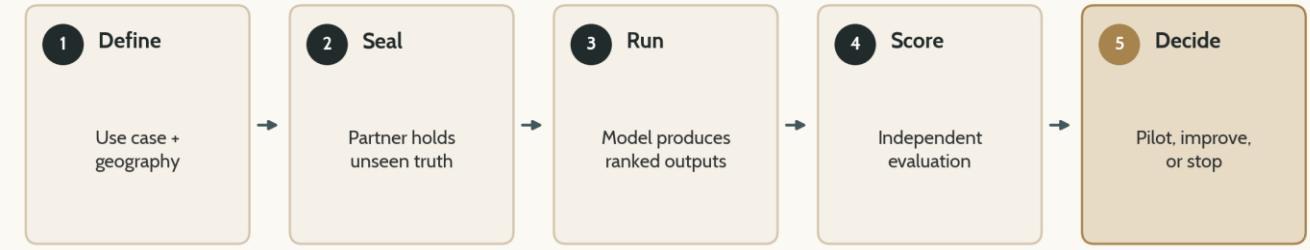
- A professional geospatial / ML engineering team
- Commercial high-resolution satellite imagery
- Expanded compute and storage capacity
- Domain experts, field access, and partner-held truth
- Independent validation governance and pilot ownership

> **The message is not "I have done everything possible."** The stronger message is that additional individual effort cannot substitute for the external data, specialist team, licensed imagery, and independent verification needed to test the system at institutional standard.

Reaching this ceiling alone reduces early technical uncertainty. It changes the next question from "can a learnable signal exist?" to "does the signal hold under independent, higher-quality, blind conditions?"

A clean blind pilot for Ibtikar

A small, well-governed pilot can turn the current internal evidence into an objective go / no-go decision.



The blind test converts strong internal evidence into evidence an investor or partner can rely on.

What the pilot should measure

- Top-ranked capture: how many known outcomes appear near the top.
- Search-area reduction: how much land can be deprioritized safely.
- False positives and false negatives.
- Ranking stability across model runs.
- Expert plausibility of the evidence explanations.
- Planning time and early screening cost avoided.

What the partner receives

- A priority map with high, medium, and low investigation zones.
- A ranked shortlist of candidate locations.
- Indicator explanations and uncertainty notes.
- Comparison of global and regional model evidence.
- An audit-ready evidence pack and evaluation summary.
- A decision record for whether to scale, refine, or stop.

Suggested Ibtikar role: help define the use case, introduce an independent data or field partner, and co-own the validation protocol. Source code and sensitive coordinates do not need to be transferred in the first-stage pilot.

A disciplined pilot is valuable even if the result is mixed. It reveals where the model generalizes, where it fails, and which commercial use case deserves the next investment.

Why this can become a business

GroundTruth AI is not selling a final answer. It is selling a better first decision before expensive field work.

FOCUS	REDUCE	EXPLAIN	RECORD
DIRECT EXPERT ATTENTION	THE AREA TO INSPECT	WHY EACH ZONE RANKED	A REVIEWABLE AUDIT TRAIL

Initial proof environment

- Archaeology and heritage provide documented locations and visible field-verification logic.
- The current evidence is strongest as a prioritization and validation workflow.

Platform expansion

- Groundwater and environmental screening.
- Natural-resource prioritization.
- Infrastructure and subsurface-risk screening.
- Each sector requires its own data, domain experts, and validation.

Defensible process moat

- Curated provenance and quality controls.
- Regional adaptation with a stable global reference.
- Leakage-aware evaluation and preserved lineage.
- Reviewable evidence outputs and protected IP.

The economic logic

Field investigation, high-resolution imagery, geophysics, drilling, and specialist review are costly. A system that reliably narrows the search area and documents why a location deserves attention can create value before the most expensive actions begin.

What an investor is backing: not a single archaeology model, but a governed geospatial decision engine that can be retrained and validated for multiple high-cost search and screening problems.

Why score comparisons alone are misleading

Published projects use different sensors, targets, geographies, and validation designs. The credible benchmark is external validation, not the highest number.

Public / internal work	What was reported	Correct investor reading
Britton et al. / Q2000	Broadscale Maya LiDAR detection across about 35,584 km^2; reported F1 0.89.	Shows the value of geographic breadth and external evaluation.
Pistola et al. / CORONA	Abu Ghraib model using historical imagery; about 90% detection accuracy and four AI-suggested sites confirmed in the field.	Shows why field verification is the strongest proof milestone.
Tadesse et al. / looting detection	PlanetScope classification of 1,943 known sites; reported F1 0.926 with spatial masking.	A strong monitoring task, but different from ranking unknown investigation zones.
Funnel Beaker visibility study	Visibility prediction using satellite and weather data; F1 up to 0.987 and AUC near 0.998 on a localized dataset.	High metrics can reflect a narrow task and do not transfer directly to another use case.
GroundTruth AI today	27 regional models; internal AUC 0.773–0.803 and Recall 0.864–0.907; controls passed and fingerprints saved. And a global model reaching 0.972 AUC	The first strong internal architecture evidence and the only model that works on massive data, independent blind proof remains the next milestone.

The correct benchmark

GroundTruth AI does not need to claim the highest metric only but it also the first project that will be published using this amount of data(more data you use more AUC you lose) all published (scientific models) are trained on very little amount of exclusive data and higher results imagery while GroundTruth is using thousands of sites and tens of thousands of data groups and(open sourced free imagery).
NOW It needs to show that its ranking improves real expert decisions on unseen data, under a locked protocol, with a transparent evidence trail.

Public-context conclusion: the most valuable next evidence is not another internal chart. It is an external pilot that can be repeated, reviewed, and believed by people who did not build the model.

Evidence ledger and disclosure boundary

The non-confidential brief explains the evidence; sensitive technical assets remain protected and can be reviewed under controlled diligence.

Evidence component	Recorded artifact	What it supports	Disclosure
Training pathway	113_train_variable_quota_regional_experts.py	How the 27 regional models were produced	Controlled review
Result summary	training_summary.json	Architecture, metric ranges, quotas, and run summary	Controlled review
Model lineage	27 saved fingerprints / hashes	Identity of the exact artifacts behind the result	Shareable summary
Evaluation hygiene	Leakage-control outcome and logs	Evidence that test controls were applied	Controlled review
Data provenance	Source and curation registry	Traceability of documented positives and controlled negatives	Non-sensitive excerpt
Model assets	Global and regional model files	Executable technical asset	Confidential / not distributed
Sensitive locations	Exact coordinates and candidate outputs	Core data and operational value	Confidential / not distributed

Non-confidential boundary: this document intentionally excludes source code, model files, exact sensitive coordinates, raw training data, credentials, proprietary feature recipes, and high-value candidate locations.

Selected public references

[1] Britton et al. (2025; journal volume 2026). “Evaluating Broadscale Deep Learning for Maya Settlement Detection in G-LiHT Lidar.” Journal of Archaeological Method and Theory. DOI: 10.1007/s10816-025-09741-5.

[2] Pistola, Orrù, Marchetti & Roccetti (2025). “AI-ming backwards: Vanishing archaeological landscapes in Mesopotamia and automatic detection of sites on CORONA imagery.” PLOS ONE 20(8): e0330419. DOI: 10.1371/journal.pone.0330419.

[3] Tadesse et al. (2026). “Satellite-Based Detection of Looted Archaeological Sites Using Machine Learning.” arXiv:2602.19608.

[4] “Machine Learning-Based Detection of Archeological Sites Using Satellite and Meteorological Data: A Case Study of Funnel Beaker Culture Tombs in Poland.” Remote Sensing 17(13):2225 (2025).

[5] GroundTruth AI internal training summary and preserved model fingerprint record, July 2026.

Proposed next conversation

A focused technical review to agree one blind professional pilot, one independent data or field partner, and a locked evaluation protocol.

GroundTruth AI | groundtruthai.tech | ali.faraj.pal@gmail.com