



GROUNDTRUTH AI

Technical Validation & Operational Evidence Dossier

Locked validation, area-wide scanning, real execution logs, score reproduction, and evidence integrity

Investor & Technical Due Diligence Edition | July 2026

Prepared by Ali Riyadh Ali Faraj | Founder & Inventor | Non-confidential edition

Scope note: this dossier reports validation and operational evidence as recorded in the cited cycles. Candidate scores are ranking outputs for expert and field verification.

1. Executive Evidence Summary

GroundTruth AI has moved beyond a single experimental model into a reproducible geospatial evidence system. The program now contains locked blind validation, matched-negative stress tests, category-specific subterranean models, two complete area-grid scans, deterministic score reproduction, protected-artifact checks, and investor-ready GIS outputs.

75% Blind portal recall 3 of 4 withheld portals	0.888 Spatial RF AUC 192 vs 1,200 matched	0.846 Country-out AUC 1,392 records	0.924 Ancient tunnel AUC Recall 0.600
6,288 South Negev rows All scored	40,401 WS_090 rows Direct real GEE	17/17 Score checks Exact matches	149/149 Integrity checks No unexplained change

Investment conclusion: the evidence supports a pilot-ready candidate-prioritisation platform with a controlled validation stack, repeatable regional scanning, and field-usable delivery formats.

The strongest technical differentiator is the evidence chain around the model: frozen artifacts, spatially realistic validation, direct identifier exclusion, hard-negative construction, independent score reproduction, de-clustering, blinded review, and SHA-256 manifests. This creates a defensible operating workflow rather than a standalone classifier.

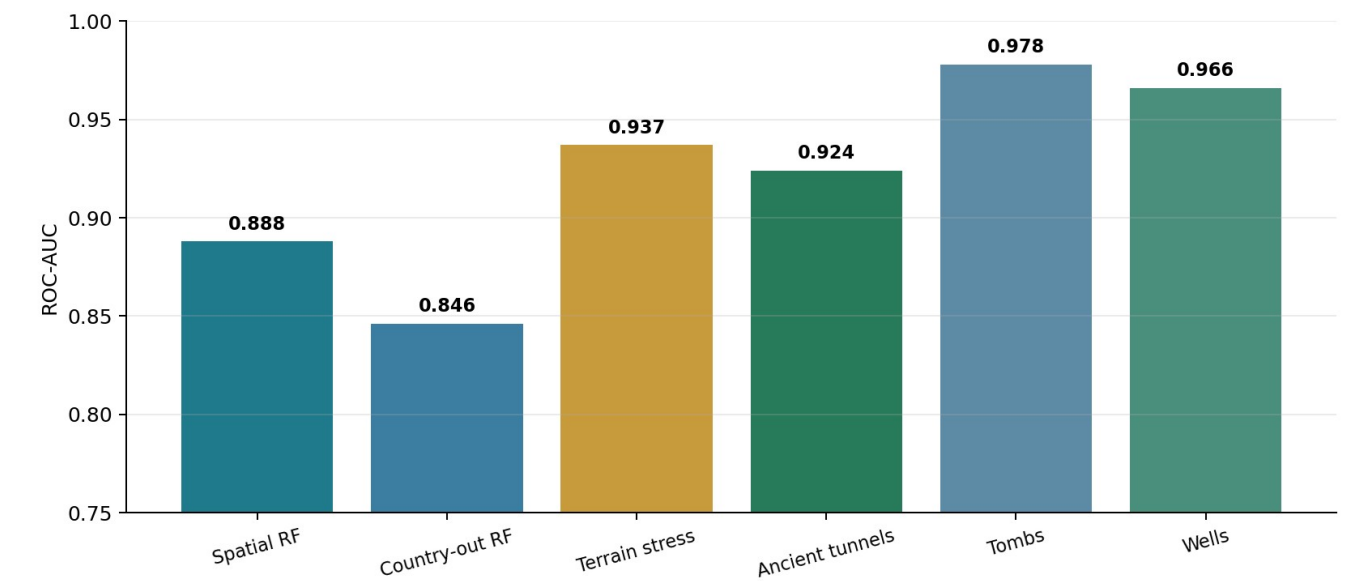


Figure 1. Locked and adversarial ROC-AUC results reported in the underlying evidence cycles.

2. Evidence Architecture and System Flow

Stage	System action	Evidence produced
1. Acquire	Sentinel-2 optical, Sentinel-1 SAR, terrain, context and temporal layers	Feature tables with source and schema records
2. Govern	Freeze models, schemas and splits; exclude direct identifiers; hash artifacts	Model contracts, manifests and integrity logs
3. Rank	Score systematic grids or candidate pools with frozen estimators	Probabilities, ranking tables and threshold views
4. Consolidate	Cluster adjacent cells and remove spatial duplicates	Operational hotspot lists and ranked AOIs

Stage	System action	Evidence produced
5. Review	Run baseline comparison, red-team flags and blind visual triage	Review packets, risk flags and escalation decisions
6. Deliver	Export CSV, GeoJSON, KML, maps, reports and hashes	Committee-ready and field-usable evidence packages

The same architecture supports different sectors because the commercial output is a ranked, auditable survey package rather than a generic heatmap. Target-specific models and verification partners can be added without changing the core evidence-control workflow.

Evidence source register

Evidence source	Role in this dossier
Investor Technical Evidence Memorandum - FINAL	Consolidated source for locked validation and operational scanning results
Investor Proof of Evidence v3 INTERIM	Tunnel-portal study, cross-class extension, and South Negev stage-one run
subterranean_class_validation_v1	Sealed split, hard-negative generation, class results, and full terminal record
live_area_scan_v1 / South Negev Phase Two	6,288-point grid, scores, maps, hotspots, and reproduction checks
V29-B / WS_090	40,401-point scan, clustering, review packets, and blind visual gate
artifact_inventory_v2 / provenance_repair_v1_2	163 model-dataset pairs, metric validation, lineage, leakage, and hashes
Website Brain Integration v1	API, Demo Mode, build, security, and integrated 6,288 + 40,401-row verification

3. Tunnel-Portal Blind Validation

The first locked cycle tested whether multi-sensor surface and terrain features could rank independently withheld civilian tunnel portals above realistic non-portal environments. Predictions were sealed before the blind labels were scored.

3/4 Blind positives recovered Frozen XGBoost threshold	100% Blind precision 0 false positives / 3 negatives	0.888 Spatial RF AUC Road/rail-matched audit	0.846 Country-out RF AUC Leave-one-country-out
---	---	---	---

Evidence layer	Scale	Result	Technical meaning
Seed blind holdout	4 portals + 3 negatives	XGB recall 0.750; precision 1.000	Withheld portal ranking demonstrated
Road-matched spatial audit	192 positives + 1,200 negatives	RF ROC-AUC 0.888	Discrimination survives matched infrastructure contexts
Leave-one-country-out	Same 1,392 records	RF ROC-AUC 0.846	Cross-country transfer evidence
Terrain stress test	192 positives + 499 steep-road negatives	RF ROC-AUC 0.937	Strong terrain-informed ranking under harder negatives

Key integrity artifacts

Artifact	SHA-256
Public registry workbook	342d3e326231f32e90d38dd1c821d12b471f3287e62b2a88f356cadedf1701dd

Artifact	SHA-256
Validated registry manifest	0720637b24bbaef56a282da2a83d7d28c1d17e9f05878d0f51dfe6ba1ab0630d
Regime B Random Forest	ecd60de32a93d2605404583b425ca24358b17568612239095f05955c0468d1b1
Regime B XGBoost	3e505a5d52e999220370dc770e4a18876d462c4f6a342b0e0935d8277d329893
Regime B CNN	01b1b0e073ad1bb8e976824e8a6dd8b314c228e41fbad6638bb97ccf0f36a4f2
Blind prediction CSV	6211ccb8fd433bed8266a4c6aee7e3f082da17d0c0042df9d47f241ffdf3a37

4. Locked Cross-Class Subterranean Validation

A second locked cycle evaluated five subterranean categories using sealed development/blind partitions, category-specific hard negatives, coordinate-only audits, feature-family ablations and a no-remediation rule after blind results were opened.

Target class	Locked result	Evidence status	Key signal
Ancient tunnels	Blind ROC-AUC 0.924; recall 0.600	Strong lead class	Spectral and SAR both contributed; no coordinate-only shortcut
Tombs / burial sites	Blind ROC-AUC 0.978	Promising	High raw signal; spatial-autocorrelation risk tracked
Caves	Blind recall 0.562; all-ready AUC 0.950	Promising	Transfer signal; more blind scale needed
Wells / cisterns	Blind ROC-AUC 0.966	Promising	Matched water-context negatives
Subterranean rooms	4 usable positives	Data expansion	Insufficient scale for a stable detector

Hard-negative construction evidence

Category	Candidate pool / negatives	Balance or control result
Tombs	3,000 OSM raw + 130 confusers; 2,906 after 250 m exclusion; 2,732 after 100 m dedup; 385 negatives	Elevation SMD 0.051; slope SMD 0.034
Caves	2,674 candidates; 407 negatives	Elevation SMD 0.185; slope SMD 0.069
Ancient tunnels	2,995 candidates; 2,503 terrain-extracted; 143 negatives	Road/rail geometry unavailable; dist_to_road left NaN rather than imputed
Wells / cisterns	2,985 candidates; 143 negatives	Elevation SMD 0.090; slope SMD 0.007
Subterranean rooms	4 positives; target 22 negatives	Data scarcity explicitly retained in final status

The locked cross-class package generated 94 hashed artifacts, including the authoritative report, machine-readable investor metrics, a complete terminal log, predictions, models, and a SHA-256 manifest.

5. South Negev Operational Evidence - Phase Two

The South Negev workstream completed a systematic regional scan using the frozen shipped Gradient Boosting artifact, its preprocessor, and the authoritative 33-feature contract. Every grid location was scored and all outputs were reconciled.

6,288	6,288	0.9962	17/17
-------	-------	--------	-------

Grid locations All accounted for 5,396 500 m hotspots Spatial shortlist	Rows scored Zero exclusions 149/149 Protected files No unexplained change	Highest score Ranked output 33 Model features Exact schema	Independent reruns Exact matches 0 Duplicate IDs Full-grid integrity
--	--	---	---

Operational element	Verified result
Grid completion	1,390 previously extracted rows + 4,898 newly extracted rows = 6,288 total
Model integrity	Model SHA-256 944e40c551fc714de66a7142be1d6bc87e1ae1d7a6db58a608778c401346979a
Preprocessor integrity	SHA-256 04b99277888e903c2e6b31ec750bf4826d0da9db02cc280268367a1b0d24bea7
Score distribution	Minimum 0.0006; median 0.0296; mean 0.1149; maximum 0.9962
Frozen operating reference	3,660 of 6,288 locations at or above 0.02
Independent verification	17 deterministic score checks passed; rank-1 difference 0.0
Outputs	Full prediction CSV/GeoJSON, KML, interactive maps, hotspot lists, reports and hashes

Phase-one reproducibility trace retained in the evidence chain

Item	Recorded value
Rows scored	2,228 feature-complete rows
Probability range	0.0008 to 0.9555; median 0.012; mean 0.048
Frozen reference	768 / 2,228 above 0.02
Rank-1 candidate	south_negev_neg_prevalence_2pct_01883
Rank-1 location / score	30.9556 N, 35.4317 E; 0.95551
Independent reproduction	Fresh artifact reload and predict_proba exact match
Map output	Real Leaflet HTML, approximately 2.9 MB

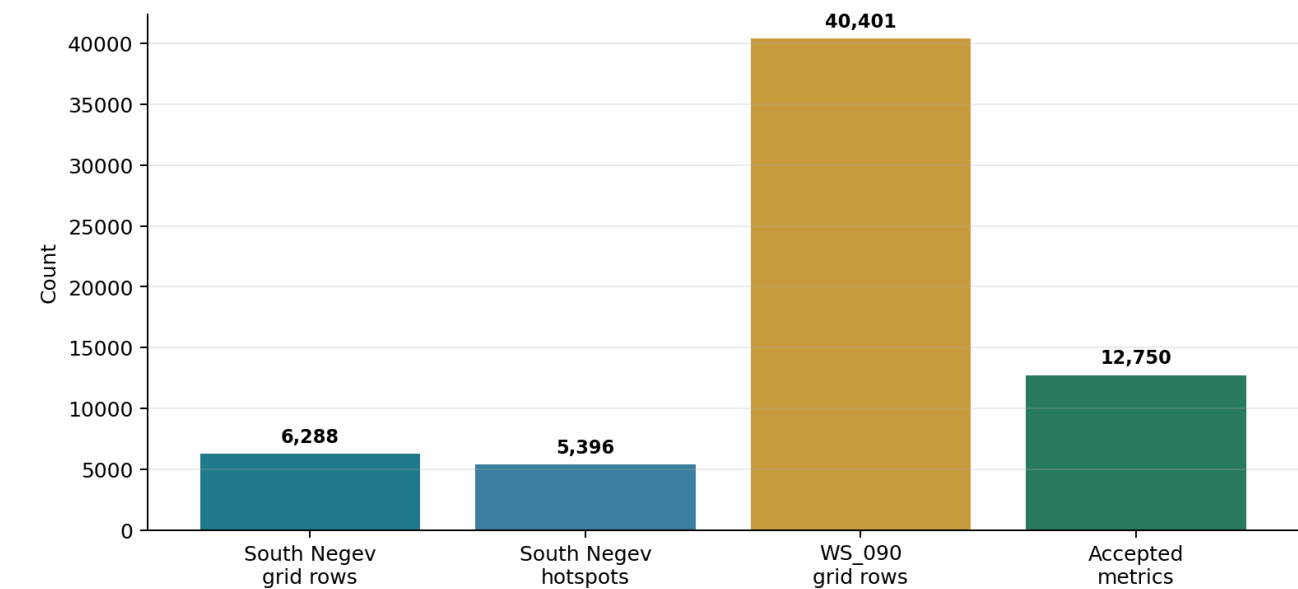


Figure 2. Operational and audit scale across the main evidence cycles.

6. V29-B / WS_090 Controlled White-Spot Scan

A second operational track scanned an approximately 400 km² white-spot region in the Eastern Jordan Basalt Margin. The completed package used real Earth Engine extraction in 25 batches and produced a full 100 m grid, model rankings, spatial clusters, blinded review packets, maps and integrity reports.

40,401 Grid rows 100 m systematic grid	25 Extraction batches Complete input package	39 Candidate clusters After 250 m dedup	100 Blind reviews 25 candidates × 4 roles
10/25/50 Review tiers Conservative / balanced / exploratory	4 GIS map outputs GeoJSON + KML	0 High-priority consensus Blind visual gate	~101 km Training separation Nearest positive

Stage	Recorded output
Geometry lock	Explicit WS_090 bounding box centered at 31.2500 N, 37.7500 E
Contamination check	No training positive in the box; nearest training positive approximately 101.4 km
Grid and extraction	40,401 real feature rows at 100 m resolution; 25 real batch files
Baseline comparison	V28 ranking compared against seven baseline views
Clustering	40 high-score groups consolidated to 39 spatial clusters
Red-team screening	0 YES, 33 MAYBE, 6 NO at the cluster screening stage
Blind panel	25 candidates, 4 reviewer roles, 100 reviews; no candidate reached average score >= 2.0
Decision gate	No field escalation from the basalt cycle; geology-specific hard-negative work recommended

This result is commercially useful because the workflow demonstrated disciplined stopping behavior in difficult geology. The system generated a complete shortlist and evidence package, then the blind review gate prevented expensive escalation when the visual evidence did not survive geological scrutiny.

7. Artifact Inventory, Provenance and Leakage Audit

The repository-wide validation program was independently audited at the model-dataset-pair level. The audit did not accept metrics merely because they appeared in reports; it reconstructed artifacts, split regimes, feature schemas, hashes and provenance wherever evidence existed.

163 Model-dataset pairs Assessed	12,750 Metrics accepted Validated extraction	60 Metrics rejected Invalid candidates	180 Ambiguous metrics Flagged
12 LOW risk pairs After correction	66 MEDIUM risk After correction	85 HIGH risk After correction	0 CRITICAL After correction

Audit control	Recorded result
Inventory v1 integrity	0 / 19 hash mismatches
Main provenance blocker	151 / 163 pairs had unresolved provenance before repair
South Negev feature firewall	0 of 14 candidate metadata/leakage fields present in the 33-feature production schema
Coordinate audit model	Latitude/longitude-only ROC-AUC 0.737-0.748; coordinates excluded from deployed inputs
Near-coordinate contamination scan	45 row pairs within 1 km; 0 exact matches below 0.5 m
Frozen baseline verification	76 frozen v2/v1/v1_1 files re-hashed; 0 mismatches
Risk decisions	0 lowered, 4 escalated, 1 unchanged in the cited repair pass
Repair package	36 files across freeze, audit, lineage, risk, logs, reports and hashes

The audit architecture is a strategic asset: it prevents unsupported fusion, separates duplicate artifacts from independent evidence, and preserves negative or blocked findings instead of converting them into marketing claims.

8. Reproducibility and Product Integrity

Control	Implementation	Evidence
Artifact freezing	Serialized models, preprocessors, schemas and thresholds are versioned and hashed	Full SHA-256 values and duplicate-copy detection
Coordinate firewall	Coordinates retained for mapping but excluded from predictive matrices	33-feature production schema; 14 metadata fields explicitly excluded
Score reproduction	Raw rows reloaded and predict_proba rerun independently	17 / 17 South Negev checks; rank-1 difference 0.0
Protected-file monitoring	Historical evidence directories checked after operational runs	149 / 149 protected-file checks passed in Phase Two
Blind handling	Predictions sealed before labels; no post-blind remediation	Tunnel and cross-class locked cycles
Spatial de-clustering	Adjacent high-score rows grouped into hotspots	6,288 rows -> 5,396 500 m hotspots; WS_090 -> 39 clusters
Abstention gate	Blind human review can reject escalation	WS_090 basalt cycle stopped before GPR preparation

Website and demo integration verification

Integration test	Result
South Negev data binding	6,288 rows and 5,396 hotspots loaded
V29-B binding	25 batches and 40,401 rows loaded with limited review package
API	PASS
Demo Mode	PASS
Production build - npm run build	PASS
General security audit	PASS
Protected technical scope	No retraining, no GEE execution, no Drive access, no model modification during website integration

9. Pilot Readiness and Investable Milestones

The technical evidence is sufficient to justify a controlled institutional pilot. The next phase should convert ranking performance and operational reproducibility into independently reviewed outcomes, customer workflow evidence and repeatable delivery economics.

Pilot workstream	Acceptance evidence	Commercial value
Partner-defined AOI	Frozen scope, target class and success criteria	Clear statement of work
Independent reference set	Partner-held sites or authorized field targets	External validation credibility
Operational scan	Complete feature extraction, ranking, clusters and maps	Demonstrates delivery speed and reliability
Expert review gate	Blinded specialist review with documented decisions	Controls escalation cost
Targeted field verification	Authorized archaeology, UAV, LiDAR, geophysics or GPR	Converts ranking to ground truth
Pilot evidence pack	Methods, outputs, logs, hashes and outcome table	Supports procurement and follow-on funding

Milestone	Target
M1 - Secure MVP	Reproducible deployment environment, access controls and customer workspace
M2 - Independent technical review	External review of split, provenance, model contract and operational logs
M3 - First institutional pilot	One heritage or infrastructure AOI processed under a frozen protocol
M4 - Field-linked outcome set	Top-ranked and control locations reviewed under an authorized protocol
M5 - Commercial package	Standardized Evidence Pack, pricing, SLA and repeatable delivery workflow

Appendix A. Selected Real Execution Log Extracts

The following excerpts are preserved as technical evidence. Formatting has been normalized for readability; values, identifiers and status lines are retained from the recorded execution summaries and terminal outputs.

A1 - Subterranean hard-negative construction (terminal excerpt)

```
[2026-07-14T16:28:45.272426+00:00] Study bbox: south=29.5830 north=33.4488 west=34.1970 east=35.8955
[2026-07-14T16:28:45.273420+00:00] === Building negatives for Ancient Tombs/Burial Sites ===
[2026-07-14T16:28:45.282776+00:00] 70 positives, target negatives ~385 (5.5x)
[2026-07-14T16:28:51.891771+00:00] attempt 1/5 FAILED: 504 Gateway Timeout - Overpass API
[2026-07-14T16:29:10.492438+00:00] attempt 2/5 FAILED: 504 Gateway Timeout - Overpass API
[2026-07-14T16:29:36.859127+00:00] OSM raw candidates: 3000
[2026-07-14T16:29:37.001779+00:00] Cross-category confuser candidates: 130
[2026-07-14T16:29:37.150659+00:00] Exclusion buffer 250.0m: 2906 kept, 224 excluded
[2026-07-14T16:29:54.399376+00:00] After spatial dedup 100m: 2732
Tombs negatives complete: 385 | SMD elevation=0.051 | SMD slope=0.034
Caves candidates recovered: 2674 | negatives=407 | SMD elevation=0.185 | slope=0.069
Ancient tunnels candidates: 2995 | terrain rows=2503 | negatives=143
dist_to_road: NaN when road/rail geometry unavailable; no all-zero substitute created
Wells/cisterns candidates: 2985 | negatives=143 | SMD elevation=0.090 | slope=0.007
Subterranean rooms: 4 positives | target negatives=22 | scarcity retained in final report
```

A2 - South Negev frozen artifact verification

```
south_negev_v1_lto10      b9ebdd3cda4cf4dc  outputs/south_negev_v1/models/south_negev_lto10_best_model.joblib
south_negev_v1_lto5       944e40c551fc714d  outputs/south_negev_v1/models/south_negev_lto5_best_model.joblib
south_negev_v1_prev2pct   9f9508e90336fc7d  outputs/south_negev_v1/models/south_negev_prevalence_2pct_best_model.joblib
south_negev_v1_recommended 944e40c551fc714d  outputs/south_negev_v1/models/south_negev_recommended_model.joblib
south_negev_v1_preprocessor 04b99277888e903c  outputs/south_negev_v1/models/south_negev_recommended_preprocessor.joblib
recommended model full SHA-256: 944e40c551fc714de66a7142be1d6bc87e1ae1d7a6db58a608778c401346979a
preprocessor full SHA-256: 04b99277888e903c2e6b31ec750bf4826d0da9db02cc280268367a1b0d24bea7
```

A3 - South Negev Phase Two completion summary

```
GRID_COMPLETION: 1390 previous + 4898 new = 6288 total
FEATURE_COMPLETE_ROWS: 6288
EXCLUDED_ROWS: 0
DUPLICATE_CANDIDATE_IDS: 0
SCORE_MIN: 0.0006
SCORE_MEDIAN: 0.0296
SCORE_MEAN: 0.1149
SCORE_MAX: 0.9962
ABOVE_FROZEN_0.02: 3660 / 6288
INDEPENDENT_SCORE_CHECKS: 17 / 17 PASS
RANK1_DIFFERENCE: 0.0
HOTSPOTS_500M: 5396
PROTECTED_FILE_CHECK: 149 / 149 PASS
OUTPUTS: predictions CSV/GeoJSON | KML | interactive maps | hotspot lists | reports | hashes
```

A4 - South Negev Stage One rank-1 reproduction

```
ROWS_SCORED: 2228
INPUT_FEATURES: 33 ordered deployed features
SCORE_DISTRIBUTION: min=0.0008 median=0.012 mean=0.048 max=0.9555
FROZEN_THRESHOLD_0.02: 768 / 2228 (34.5%)
RANK1_ID: south_negev_neg_prevalence_2pct_01883
RANK1_COORDS: 30.9556 N, 35.4317 E
SAVED_PROBABILITY: 0.95551
RECOMPUTED_PROBABILITY: 0.95551
ABSOLUTE_DIFFERENCE: 0.0
MAP: outputs/live_inference_demo_v1/maps/all_predictions_map.html (~2.9 MB)
```

A5 - V29-B / WS_090 output inventory

```
REGION: WS_090 - Eastern Jordan Basalt Margin
CENTER: 31.2500 N, 37.7500 E
AREA: approximately 400 km2
NEAREST_TRAINING_POSITIVE: approximately 101.4 km
GRID_ROWS: 40401 @ 100m
REAL_EXTRACTION_BATCHES: 25
CANDIDATE_GROUPS: 40 -> 39 after 250m clustering
RED_TEAM_CLUSTER_FLAGS: 0 YES | 33 MAYBE | 6 NO
REVIEW_PACKAGES: top10 conservative | top25 balanced | top50 exploratory
BLIND_PANEL: 25 candidates x 4 roles = 100 reviews
PANEL_RESULT: 0 candidates average >= 2.0 | 0 scores of 3
GIS_OUTPUTS: all-points GeoJSON | cluster GeoJSON | top10 KML | top25 KML
DECISION: stop field escalation in basalt-dominated terrain; add geology-specific hard negatives
```

A6 - Artifact Inventory v2 / provenance repair summary

```
MODEL_DATASET_PAIRS_ASSESSED: 163
```

METRICS_ACCEPTED: 12750
METRIC_CANDIDATES_REJECTED: 60
AMBIGUOUS_METRICS_FLAGGED: 180
LEAKAGE_RISK: LOW=12 MEDIUM=66 HIGH=85 CRITICAL=0
UNRESOLVED_PROVENANCE_INITIAL: 151 / 163 pairs
V1_INTEGRITY: 0 / 19 hash mismatches
NEAR_COORDINATE_SCAN: 45 pairs within 1000m | 0 exact <0.5m
PRODUCTION_FEATURE_FIREWALL: 0 / 14 candidate metadata fields present
FROZEN_BASELINES_REHASHED: 76 files | 0 mismatches
RISK_CHANGES_IN_REPAIR_PASS: 0 lowered | 4 escalated | 1 unchanged
REPAIR_OUTPUT_MANIFEST: 36 files

A7 - Website / demo integration verification
SOUTH_NEGEV_DATA: 6288 rows | 5396 hotspots
V29B_DATA: 25 batches | 40401 rows | limited review package
API: PASS
DEMO_MODE: PASS
NPM_RUN_BUILD: PASS
GENERAL_SECURITY_AUDIT: PASS
NO_TRAINING: TRUE
NO_GEE_EXECUTION: TRUE
NO_DRIVE_ACCESS: TRUE
NO_MODEL_MODIFICATION: TRUE

Appendix B. Evidence Package and Data-Room Index

Evidence family	Representative artifacts
Tunnel portal validation	Registry workbook; validated manifest; sealed predictions; RF/XGB/CNN models; spatial and country-out audit tables; hashes
Cross-class subterranean validation	FINAL_SUBTERRANEAN_CLASS_VALIDATION_v1.md; investor_metrics_summary.json; full terminal log; 94-artifact SHA-256 manifest
South Negev stage one	LIVE_INFERENCE_DEMO_REPORT_v1.md; ranked predictions; all_predictions_map.html; model/preprocessor hashes
South Negev phase two	Full 6,288-row prediction table; hotspot lists; GeoJSON/KML/maps; 17 reproduction checks; 149-file integrity report
V29-B / WS_090	25 real feature batches; 40,401 predictions; cluster tables; risk flags; blind/ranked review packets; GeoJSON/KML
Artifact inventory and provenance repair	Validated inventories; risk tables; lineage reconstruction; contamination audit; feature firewall report; hash manifests
Investor delivery layer	Technical dossier; detailed implementation proposal; business plan; pitch deck; secure demo

Recommended controlled data-room sequence

- Level 1 - Investor package: pitch deck, investor brief, this technical dossier, implementation proposal and financial plan.
- Level 2 - Technical diligence: authoritative reports, machine-readable metric summaries, model-card style contracts, selected logs and hash manifests.
- Level 3 - Restricted review: source-level manifests, protected coordinate tables, reproducibility scripts and selected model artifacts under NDA.
- Level 4 - Partner pilot workspace: partner-defined AOI, frozen protocol, review outputs and field-linked outcome table.

No log in this dossier was invented. Values were taken from the existing validation papers, execution summaries, terminal excerpts, model hashes, artifact inventories and integration reports listed in the source register.