Evidence-Grounded Plan-to-Field Telecom Anomaly Verification with Tool-Using Multimodal Agents

Mahdi Setayesh
Microsoft, Industry Solutions Engineering (ISE)
Redmond, USA
msetayesh@microsoft.com

Paul Butler
Microsoft, Industry Solutions Engineering (ISE)
Redmond, USA
paulbutler@microsoft.com

Maitreyi Nair
Microsoft, Industry Solutions Engineering (ISE)
Redmond, USA
maitreyi.nair@microsoft.com

Kelsey Huebner
Microsoft, Industry Solutions Engineering (ISE)
Redmond, USA
kehueb@microsoft.com

James Lee
Microsoft, Industry Solutions Engineering (ISE)
Redmond, USA
james.lee@microsoft.com

Shane Patterson
Microsoft, Industry Solutions Engineering (ISE)
Redmond, USA
shane.patterson@microsoft.com

ABSTRACT

Construction design documents for telecom towers frequently drift from what is ultimately installed in the field: antennas can be added or removed, equipment variants change, and physical dimensions differ from the specified parts. Detecting these plan-to-field discrepancies is important for reliability and auditability, yet the current workflow remains largely manual, requiring engineers to reconcile schematics with drone imagery site by site.

We introduce a **tool-using multimodal agent** for scalable plan-to-field anomaly verification. Given design documents and post-construction drone photos, the agent plans an inspection workflow, extracts expected configuration from documents, and applies vision tooling (detection, segmentation) together with depth-based measurement to produce evidence-grounded comparisons. Rather than treating verification as a single fixed computer vision pipeline, the agent adaptively selects tools and re-checks evidence when views are ambiguous or incomplete, and it outputs an audit-style report that cites both document and image artifacts. Our contribution is *not* a new perception model, but a verification-oriented agentic workflow optimized for auditable decisions in high-consequence infrastructure contexts.

We evaluate the approach on real tower audits, emphasizing two common and operationally meaningful checks: antenna count and antenna physical dimensions (with orientation supported as an optional rule). The results show the agent can surface mismatches with traceable supporting evidence, and the modular design supports extending coverage to additional anomaly types as new tools become available. Source code: https://github.com/azure-samples/multimodal-agent-telecom-anomaly-detection.

KEYWORDS

Anomaly Detection, Multimodal Agents, Tool Use, Telecom Infrastructure, Vision-Language

2nd Workshop on AI for Critical National Infrastructure (AI4CNI-26) @ AAMAS 2026, F.L. da Silva, W. Su, R. Glatt, C. Schroeder de Witt, May 25 – 29, 2026, Paphos, Cyprus. © 2026 Copyright held by the author(s).

1 INTRODUCTION

Telecom tower audits routinely uncover gaps between what was planned and what was ultimately installed. Common issues include missing or extra antennas, equipment swaps, and mismatched physical dimensions. These discrepancies matter for capacity planning, safety, and operational reliability, yet verifying them still relies heavily on manual review: engineers compare construction design (CD) artifacts against drone imagery, triangulate ambiguous viewpoints, and document findings. This process is time-consuming, exposes teams to risk (especially when follow-up site visits are needed), and does not scale to large fleets.

This manual inspection is **slow, risky, and not scalable**, and it is prone to human error. In practice, an engineer may still need to examine drone photos with design schematics in hand, checking each item one by one.

Automation is challenging because the "ground truth" is split across modalities. Design documents encode intended configuration, sectors, and part specifications, while field evidence arrives as heterogeneous image sets with occlusions, inconsistent coverage, and limited metadata. Conventional computer vision pipelines can address a narrow set of visual tasks (e.g., detect antennas or count objects), but they often generalize poorly when tower layouts vary or when a new anomaly category is introduced. Several automated approaches have emerged to bridge the gap between design and reality. Computer vision (CV) techniques on drone imagery can detect certain issues quickly: for instance, an object detection model might count antennas or flag physical damage. Such vision-based methods work well for the specific anomalies they are trained on. However, they struggle with new or complex issues outside their training data. Each novel anomaly might require re-training or new algorithms, which is impractical given the *hundreds* of anomaly types encountered in tower audits. Neither the traditional CV pipelines nor the digital twin methods fully exploit the rich information in design documents; they largely treat the problem as "find differences in images," without deeply understanding what the design specifies. At the other extreme, digital-twin style approaches reconstruct a 3D model from images or LiDAR and compare geometry against plans [1, 7, 18]. These methods can be powerful, but producing

accurate reconstructions at scale is expensive and brittle, and the comparison still requires interpreting what the documents specify.

Prior tower-focused work explores 3D reconstruction and point-cloud analysis (e.g., SfM-based modeling [2] and 3D tower asset classification [11]), as well as UAV LiDAR inspection in closely related infrastructure domains [9, 12, 19]. However, in operational audits the bottleneck is often not a single perception primitive, but end-to-end reasoning under incomplete evidence: choosing which views to trust, mapping field observations to design entities, and stating conclusions with traceability.

In practice, towers exhibit a **wide variety of anomaly types** and field conditions, and the available evidence is rarely ideal. An antenna that appears "missing" in one photo may be visible in another; a sector-facing view may be partially occluded by nearby equipment; and construction drawings may be scanned, annotated, or organized inconsistently. In these settings, a useful system must do more than produce a one-shot prediction: it must plan an inspection, re-check evidence when uncertain, and explicitly surface evidence gaps when a determination cannot be made reliably. A rigid automation pipeline might flag a missing antenna but fail to reason that the antenna could simply be obscured in one photo and visible in another. Many verification tasks require adaptive reasoning: under controlled lab conditions (a "studio" scenario), one could verify an antenna's orientation easily; in the field, it might require examining multiple images and correlating with design info to conclude an antenna is mis-angled. We therefore treat verification as a sequential, rule-driven process where an agent reads criteria, retrieves the required sheets and views, and iterates when ambiguity is detected, producing an audit-style report with stable citations to both document artifacts and image-derived evidence.

Large language model (LLM) agents provide a complementary approach: an LLM can act as a controller that sequences tools and adapts its workflow when information is missing. ReAct [17] demonstrates interleaving reasoning with actions, and multimodal agent work extends this concept to jointly operate over text and images [4, 8]. We apply these ideas to telecom plan-to-field verification, where the task naturally requires both document understanding and visual inspection.

We propose a **tool-using multimodal agent** that verifies anomaly rules by cross-referencing design intent with field evidence. The agent uses non-vision tools to parse and index CD documents and retrieve expected configuration, and vision tools to interpret drone imagery (detection, segmentation, and depth-based measurement). Our contribution is *not* a new perception model, but a verification-oriented agentic workflow optimized for auditable decisions: it orchestrates tools to gather evidence, re-checks ambiguous cases, and produces a report with citations rather than a black-box prediction. Rather than emitting a single scalar prediction, the agent produces an evidence-grounded, audit-style report: it quotes the relevant criteria from the design artifacts, links them to measured or detected field evidence, and states pass/fail outcomes per criterion. When evidence is insufficient (e.g., occlusions or missing viewpoints), the agent can surface the gap explicitly and request alternative images instead of guessing.

Inspired by these developments, we propose applying an **agentic, multimodal approach** to telecom tower anomaly verification. To the best of our knowledge, no prior work in the telecom domain has utilized an AI agent that reads design documents and inspects images in tandem to find discrepancies.

This paper emphasizes two high-impact checks that combine perception with numeric comparison: antenna count and antenna physical dimensions. Orientation verification is also supported as an optional rule but is not the primary experimental focus. Across real towers, we show the agent can identify mismatches such as missing antennas and dimension deviations while maintaining traceability via cited document excerpts and image-derived measurements.

These serve as test cases for our agent because they are common sources of errors and require multimodal reasoning (counting and measuring from images, comparing with numeric specifications). In our experiments on real-world towers, the agent successfully identifies anomalies in these aspects that were difficult to catch with prior methods. For example, it caught cases where a sector had an antenna missing that the design listed (the agent knew to look for it and double-checked multiple images) and cases where an antenna's length did not match the model specified in the inventory (the agent read the model's spec sheet and compared it to measurements from the image).

From a critical national infrastructure perspective, the goal is not only automation but *assurance*. We therefore emphasize evidence-grounded decisions: the agent ties each pass/fail outcome to cited design artifacts and image-derived evidence, and it records explicit evidence gaps (e.g., occlusion or missing viewpoints) instead of guessing. This report-first, tool-traced workflow supports (i) *resilience* by enabling targeted re-inspection when evidence is missing, (ii) *regulatory and internal compliance* workflows by providing artifact-level provenance for each decision, and (iii) *risk mitigation* by discouraging overconfident remediation recommendations and making uncertainty explicit.

The remainder of the paper formalizes plan-to-field verification, describes the agent architecture and tool suite, and presents experiments and qualitative analyses illustrating how an agentic approach improves adaptability and explainability in telecom audits.

2 PROBLEM SETUP: PLAN-TO-FIELD ANOMALY VERIFICATION

We study rule-based verification where the intended (planned) configuration for a site s is compared against post-construction field evidence. Inputs include construction design (CD) documents (PDF sheets), optional equipment schedules, and drone images captured after construction with available EXIF metadata. The system outputs a *report-only* assessment: a pass/fail-style decision for each rule, a rationale, and explicit citations to the underlying evidence.

In practice, telecom audits cover a broader space than plan-vs-as-built equipment matching. Many anomalies are compliance- and standards-driven, such as missing or incorrect safety signage (e.g., RF exposure warnings), noncompliant labeling/markings, or missing zoning/permit information that must be verified against policies and regulations. Such checks can also be expressed as anomaly rules when the relevant evidence is available (documents, imagery, or structured site records). In this paper, however, we primarily focus on plan-to-field comparisons of tower antennas by comparing as-designed specifications to as-built evidence for antenna count and physical dimensions.

A key difficulty is establishing correspondence between plan entities and field views. Design documents specify sectors and azimuths, but the "best" post-construction image for a given sector is not known in advance and must be retrieved from a large image set (e.g., using azimuth-guided selection).

3 PROPOSED METHOD: MULTIMODAL TOOL-USING AGENTS

3.1 Agent Architecture

We structure plan-to-field verification as a two-agent workflow with complementary roles (Figure 1).

Detector Agent: A configurable tool-using agent that executes a rule workflow by orchestrating domain tools. It loads the anomaly rule and site metadata, indexes construction drawings, selects sector-relevant imagery, detects antennas, and (when required) estimates physical dimensions via depth-based 3D measurements.

The detector returns (1) `anomaly_detected` and (2) a prescribed Markdown `report` organized into seven sections (Summary, Scope & Context, Standards & Criteria, Evidence Catalog, Analysis, Findings, Report Summary). This format supports auditability and evaluation; Appendix C provides an example.

Reviewer Agent: Checks that the detector followed the required workflow and enforced a strict *report-only* policy. The reviewer has no tools and sees only the detector's structured output. Validation covers seven dimensions:

(1) *Data scoping*: Site and rule identifiers are present and valid.
(2) *Rule grounding*: Detection criteria from the anomaly rule are cited.
(3) *Evidence retrieval*: All referenced files, images, and measurements exist.
(4) *Analysis execution*: Evidence was actually examined (not just listed).
(5) *Criteria comparison*: Each criterion has a pass/fail determination.
(6) *Findings synthesis*: Discrepancies include severity and rationale.
(7) *Report summary*: Generation metadata present; no remediation recommendations.

The reviewer returns `decision` (`COMPLETE` or `MISSING_STEPS`) and, when needed, `missing_steps` guidance to drive a re-run. Any remediation or "what to fix on site" recommendations trigger rejection. The loop runs for up to n rounds.

3.2 Tool Suite

The detector agent uses a small set of specialized tools:

(1) **Anomaly Information Retrieval:** loads the anomaly rule, including criteria, comparison logic, severity thresholds, and required evidence artifacts.
(2) **Site Information Retrieval:** resolves site metadata such as CD document paths, image folder structure by construction phase, and tower GPS coordinates.
(3) **Construction Drawing Indexing:** categorizes CD PDF pages (title sheet, elevation, final antenna plan, equipment schedule, etc.) for targeted retrieval.
(4) **OCR-Assisted Vision Analysis:** analyzes rendered pages with optional OCR. Structured prompts extract GPS (title sheet), antenna centerline levels (elevation), and per-sector counts and azimuths (final antenna plan), plus any additional rule-required fields.
(5) **Find Best View:** queries a spatial index of drone images to select a post-construction view whose viewing direction best aligns with a target sector azimuth, reducing occlusion.
(6) **Depth Estimate and 3D Insight Information:** detects antennas and enriches each detection with oriented boxes, monocular depth-based range estimates, physical dimension estimates, and 3D visualization. Section 4 details this tool.

3.3 Antenna Count Discrepancy Detection (ANOM1)

The antenna count discrepancy rule detects mismatches between the number of antennas specified in **final** construction drawings and the number visible in post-construction drone imagery. This is a common audit check: confirming that planned equipment appears in the field.

Each anomaly rule encodes both (i) comparison criteria and (ii) a *procedural workflow*: which evidence to read, what fields to extract, required tool calls, and pass/fail logic. This separates domain knowledge (rule authoring) from execution capabilities (tools), allowing domain experts to add rules without changing code. At runtime, the detector loads the rule and executes its steps.

ANOM1 is implemented as an auditable workflow that guides how the detector operates:

(1) Load the rule criteria and the site manifest.
(2) Index CD pages and locate required sheets (if missing, record an evidence gap).
(3) If tower GPS is absent from site metadata, extract it from the title sheet (OCR + structured prompt).
(4) From the elevation sheet, extract antenna centerline levels (ACLs) and identify any proposed ACL updates.
(5) From the **FINAL** antenna plan diagram, extract planned antenna counts and azimuths per updated ACL and sector. Using OCR-assisted vision-language analysis, the detector emits a structured representation with sector azimuth (degrees) and planned antenna count. Only the "FINAL" diagram is used; existing and proposed-removal diagrams are ignored.
(6) For each target sector azimuth, retrieve the post-construction drone image whose viewing angle best matches that bearing. Detect panel antennas in the selected image and filter detections using a calibrated confidence threshold. The detection count is treated as the actual count for that sector.
(7) Compare planned vs. actual counts per sector (and per ACL when multiple updated ACLs exist) and assign severity. *Critical*: absolute planned-vs-actual difference ≥ 2 for any ACL/sector pair, or mismatches across multiple sectors on the same ACL. *Major*: absolute difference of 1 for an ACL/sector pair. *Minor*: indeterminate count due to evidence quality (logged as an evidence gap rather than treated as a pass). On model error, retry once; if still failing, record an evidence gap.
(8) Generate the report with citations and an explicit no-notification summary.

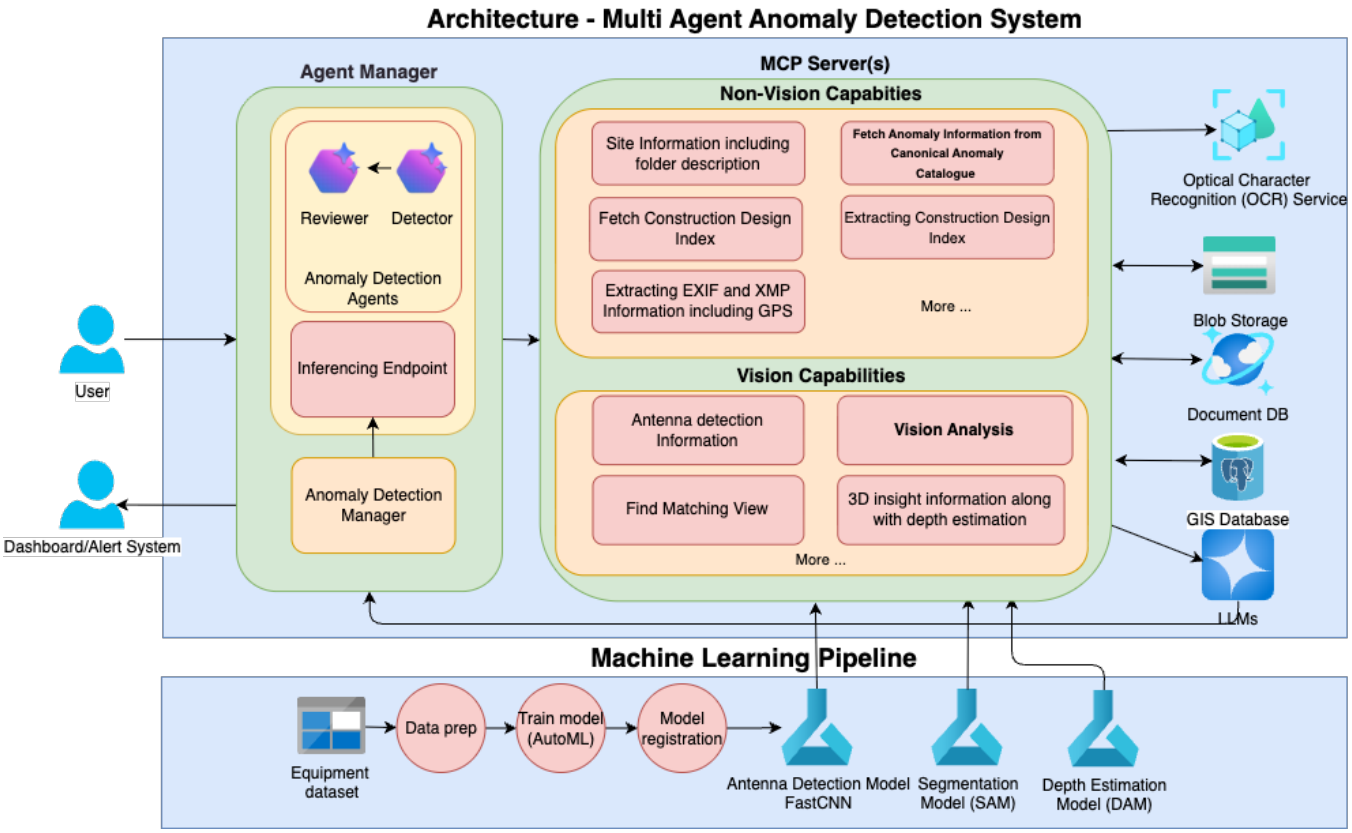


Figure 1: Overview of the proposed plan-to-field verification workflow. The Detector agent orchestrates multimodal tools (document indexing, vision, measurement) to produce an evidence-grounded report, and a tool-free Reviewer agent validates completeness while enforcing the report-only policy.

3.4 Antenna Dimension Verification (ANOM2)

The antenna physical design verification rule checks whether antenna dimensions in **final** construction drawings match dimensions estimated from post-construction drone imagery. Unlike ANOM1, this rule is measurement-driven, so the detector uses 3D insight to estimate antenna height/width and range for each detected panel antenna:

(1) Reuse ANOM1 steps 1–4 (rule/site loading, CD indexing, GPS extraction if needed, ACL extraction).
(2) From the **FINAL** antenna plan diagram, extract azimuths per ACL/sector; from the equipment schedule, extract planned antenna physical dimensions.
(3) For each target sector azimuth, retrieve the best-aligned post-construction drone image; then detect antennas and estimate dimensions (height/width) and distance using depth-based 3D analysis.
(4) Associate detections with planned specifications and compare planned vs. estimated dimensions per ACL/sector. Mark a *mismatch* if any dimension differs by $> 10\%$. Severity: *Critical* for mismatch $> 20\%$; *Major* for 10–20%; *Minor* for $< 10\%$; and *Evidence Gap* when detection fails, distance estimation is unavailable, or image quality is insufficient. On model error, retry once; if still failing, record an evidence gap.
(5) Produce a report with citations, 3D visualization, and an explicit no-notification summary for human review.

4 ANTENNA MEASUREMENT AND 3D INSIGHT VIA DEPTH ESTIMATION

Some anomaly rules (e.g., ANOM2) require estimating physical dimensions rather than counting objects. To support this, we implement a four-stage pipeline that extracts 3D-relevant cues from monocular drone imagery. We build on Nielsen [10] and extend the approach with segmentation and metric depth estimation to produce oriented measurements suitable for antenna verification.

Concretely, each image is processed by: (1) antenna detection, (2) instance segmentation, (3) oriented bounding box (OBB) fitting, and (4) metric depth estimation combined with camera intrinsics to map pixel extents to physical dimensions.

4.1 Stage 1: Antenna Detection

We train a custom detector using AutoML with a Faster R-CNN backbone [13]. Training uses tower imagery labeled with axis-aligned bounding boxes (AABB) for panel antennas. At inference, the model produces candidate boxes with confidence scores. We adopt this approach because direct vision-language counting (GPT-5, Mistral, Llama, Gemini) is unreliable in our setting, exhibiting both low accuracy and high variance.

4.2 Stage 2: Instance Segmentation

For each detected box, we compute a pixel-level mask using the Segment Anything Model (SAM; 6). The detection box is used as a spatial prompt,

Algorithm 1 Oriented Bounding Box Fitting

Convert the segmentation mask to an 8-bit binary image to enable contour extraction.
Extract contours from the binary mask using OpenCV.
If no contours are found: return `None` (no valid OBB).
Select the largest contour by area.
Compute the convex hull of the largest contour.
Simplify the hull polygon using [3] approximation (e.g., $\epsilon = 0.02 \cdot \texttt{arcLength}$).
If the simplified polygon has four vertices: use them as initial corners.
Otherwise, compute four extreme corners from hull points using coordinate sums/differences.
Identify four extreme corner points (top-left, top-right, bottom-right, bottom-left) using the sum and difference of coordinates.
Compute the corner centroid and sort corners by polar angle around the centroid to ensure a consistent clockwise order.
Apply the parallelogram constraint: adjust corner positions so that the diagonals bisect each other (i.e., $\frac{p_0+p_2}{2} = \frac{p_1+p_3}{2}$), minimizing total displacement from the original points.
Output the four adjusted corners (p_0, p_1, p_2, p_3) as the oriented bounding box.

and SAM returns a binary mask approximating the antenna silhouette. Compared to an AABB, the mask better captures shape and orientation, especially under oblique viewpoints.

4.3 Stage 3: Oriented Bounding Box Fitting

Panel antennas are approximately rectangular in 3D. Under perspective projection they appear as quadrilaterals; for simplicity we fit a parallelogram (rather than a fully general quadrilateral, often a trapezoid). We fit a parallelogram to each segmentation mask using Algorithm 1.

4.4 Stage 4: Metric Depth Estimation and Dimension Information

We estimate per-pixel depth using Depth Anything V2 [16], a metric monocular depth model based on a Vision Transformer (ViT) with a dense prediction transformer (DPT) decoder. The model is calibrated on KITTI outdoor data and outputs depth in meters. For each antenna mask we compute depth statistics (median, mean, min, max, std). Physical dimensions are computed via the pinhole projection (Algorithm 2), using OBB edge lengths as pixel extents.

5 EXPERIMENTS

5.1 Experimental Setup

We evaluate the system from an end-to-end verification perspective and complement quantitative metrics with qualitative inspection of agent-generated reports and intermediate visual artifacts. While we record agent behavior signals (tool-use patterns, robustness, and cost/latency), they are not the primary focus here. Our evaluation targets: (1) how iterative improvements to ANOM1 increase plan-to-field accuracy, and (2) how well the 3D insight pipeline supports antenna dimension verification for ANOM2. Because publicly available datasets do not provide tower-scale collections with the capture metadata needed for metric depth validation, we report qualitative analysis and limited numerical checks for ANOM2.

5.1.1 Operational Signals (Robustness, Cost, and Latency). In addition to task-focused quality metrics, we log operational signals that influence real audit throughput. For each run, the system captures tool invocation traces (which tools were called and how often), end-to-end runtime, and retry/failure events (e.g., OCR parse failures or vision-model errors). These

Algorithm 2 Stage 4: Metric depth estimation and dimension calculation

Input: Image I; detections with segmentation masks M_i and OBB corners $(p_0^i, p_1^i, p_2^i, p_3^i)$
Output: Per-detection distance statistics and physical dimensions (width/height)
Estimate a dense metric depth map Z (in meters) for the full image using Depth Anything V2.
For each detection i:
  Extract depth values $\mathcal{Z}_i \leftarrow \{Z(x, y) \mid (x, y) \in M_i\}$.
  If $|\mathcal{Z}_i| = 0$: mark an evidence gap and continue.
  Compute depth statistics: median D_i, mean, min, max, and standard deviation of $\mathcal{Z}_i$.
  Read focal length f (in pixels) from EXIF metadata; if unavailable, mark an evidence gap and continue.
  Compute pixel width $w_i^{\text{px}} \leftarrow (\|p_1^i - p_0^i\| + \|p_2^i - p_3^i\|)/2$.
  Compute pixel height $h_i^{\text{px}} \leftarrow (\|p_2^i - p_1^i\| + \|p_3^i - p_0^i\|)/2$.
  Convert pixels to physical size using the pinhole projection: $d_{\text{physical}} = \frac{d_{\text{pixels}} \times D_i}{f}$, where $d_{\text{pixels}} \in \{w_i^{\text{px}}, h_i^{\text{px}}\}$.
  Compute physical width $w_i \leftarrow (w_i^{\text{px}} \cdot D_i)/f$ and height $h_i \leftarrow (h_i^{\text{px}} \cdot D_i)/f$.
  Convert meters to feet for reporting (using 1 m = 3.28084 ft).

signals help identify bottlenecks where additional evidence gathering improves correctness but increases latency, and they make it possible to localize brittle steps when results degrade. We also probe robustness by repeating runs on the same site because stochastic multimodal prompting can yield inconsistent outputs, especially when imagery is ambiguous. While a full cost and latency study is outside the scope of this paper, we include run metadata in the generated reports and preserve execution traces to support systematic evaluation and ablations.

5.1.2 Datasets. We use two non-overlapping datasets:

(i) a labeled image set for training and evaluating the antenna detector (61 training images, 91 test images) drawn from post-construction drone audits for a US telecom provider. The imagery and site materials are proprietary and cannot be released; we therefore report dataset sizes, splits, and evaluation protocols and release the full implementation and experiment configurations (see the URL in the abstract). While this limits direct replication on the identical sites, the proposed contribution is the verification-oriented, rule-driven agentic workflow and its evidence-grounded reporting interface. In particular, the approach transfers to other operators and regions as long as (i) design intent is available in document form (PDF drawings, equipment schedules, or similar) and (ii) field evidence is available as post-construction imagery. To support generalizability claims despite data constraints, we release the end-to-end pipeline structure, prompts, and configuration files, and we include public-image qualitative visualizations; future work will extend validation using simulated perturbations (occlusion, missing views, metadata dropouts) and broader ablations over alternative detection and depth tools. For qualitative visualizations, we also use public tower imagery from the Roboflow Universe telecom antenna detection dataset [14]. Public datasets typically lack tower-level drone collections with consistent EXIF metadata required for azimuth-guided retrieval and metric measurements, so quantitative analyses rely on the proprietary imagery.

(ii) For end-to-end ANOM1 evaluation, we curated 39 tower sites. Each site includes multi-page construction drawings (title sheet, tower elevation, final antenna plan) and a full post-construction drone image collection stored in blob storage and indexed in PostgresGIS using EXIF metadata. We manually curated (a) planned antenna counts from CDs and (b) actual antenna counts from drone imagery at the site-ring level.

Figure 2: Example images from the public Roboflow Universe telecom antenna detection dataset.

To establish ground truth for *planned* counts, we manually read the **FINAL** antenna plan diagram and any co-located antenna information tables to record per-sector azimuths and expected panel counts. To establish ground truth for *actual* counts, we select a human-validated subset of post-construction views that best face each sector and count visible panel antennas in those images. When imagery is insufficient for a confident determination (e.g., heavy occlusion or missing sector coverage), we record an evidence limitation instead of forcing a count. This protocol mirrors operational audits and supports evaluating whether the agent (i) extracts the correct planned structure, (ii) retrieves an appropriate sector-facing view, and (iii) produces a traceable conclusion grounded in the cited evidence.

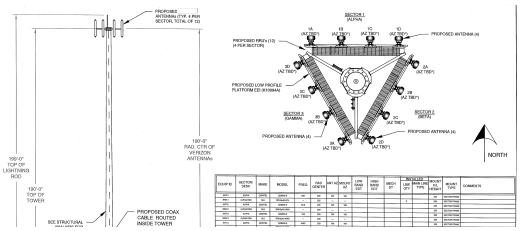


Figure 3: Example antenna plan diagram from a construction drawing PDF page. The detector agent extracts planned antenna counts and per-sector azimuths from these diagrams using OCR-assisted MLLM-based vision analysis (source: [5]).

5.1.3 Evaluation Metrics. To evaluate the end-to-end agentic workflow, we use the following task-focused metrics:

Count Accuracy. Fraction of sites where the predicted count exactly matches ground truth. For a dataset of N sites with ground-truth C_i and predicted $\hat{C}_i$ for site i, we report

$$\text{Acc} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\{\hat{C}_i = C_i\}. \tag{1}$$

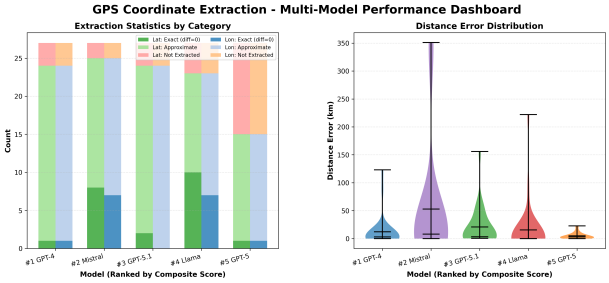
For planned count accuracy, C_i denotes the ground-truth planned antenna count and $\hat{C}_i$ the agent-extracted planned antenna count. For actual count accuracy, C_i denotes the ground-truth actual antenna count and $\hat{C}_i$ the agent-produced actual antenna count.

Image Retrieval Quality. For azimuth-guided view selection, we report the *absolute distance (feet)* between selected and ground-truth image locations (mean and median).

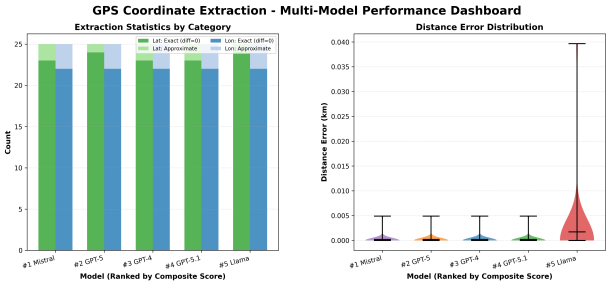
GPS Extraction Accuracy. Great-circle distance between extracted and ground-truth coordinates via the Haversine formula [15].

Evidence Gap Rate. Fraction of site-rule checks that terminate with an explicit evidence gap (e.g., missing sheets, unreadable OCR, or insufficient sector-facing viewpoints). This metric captures a key operational distinction between (i) a confident pass/fail and (ii) a workflow that correctly identifies when additional evidence collection is needed.

Reviewer Completeness Rate. Fraction of runs where the Reviewer returns `COMPLETE` without requesting missing steps. This approximates how often the end-to-end workflow produces a policy-compliant report on the first attempt and helps localize failure modes in rule execution or evidence citation.



(a) Direct vision-based extraction (no OCR).



(b) OCR-assisted extraction.

Figure 4: GPS coordinate extraction on CD title sheets. OCR-assisted extraction substantially improves reliability and accuracy relative to direct vision-based parsing.

5.1.4 Information Extraction from Construction Drawings. The agent must extract structured fields from construction drawings (CDs) to drive downstream tools. We therefore evaluate GPS (latitude/longitude) extraction from CD title sheets, because tower GPS directly affects azimuth-guided view selection and end-to-end accuracy. We compared five MLLMs (GPT-4, GPT-5, GPT-5.1, Llama, Mistral). Without OCR, direct vision-based extraction is inconsistent (misses and outliers), with Llama yielding the best median accuracy when it succeeds. With OCR preprocessing, all models become reliable on the evaluated set and GPT-5 achieves the best mean accuracy. Accordingly, we use OCR-assisted extraction when site metadata lacks GPS, and we use GPT-5 for the remaining experiments.

5.2 Evaluation of ANOM1 Antenna Count Discrepancy Detection

We evaluate ANOM1 across four progressively improved configurations:

RUN01: Baseline detector using GPT-5 for planned-count extraction (from diagrams) and actual-count estimation (from drone images via vision-language prompting).

RUN02: Same as RUN01, but planned-count extraction is augmented with OCR.

RUN03: As RUN02, with prompt improvements to extract from both the antenna table and the plan diagram.

RUN04: Detector equipped with a trained antenna detector for actual counts using a calibrated confidence threshold (0.9), while still using GPT-5 with OCR for planned count extraction.

Score-threshold calibration details appear in Appendix A. Table 1 summarizes ANOM1 accuracy across iterations. Planned-count accuracy improves

Table 1: End-to-end ANOM1 results across iterations. *Planned Antenna Accuracy* measures extraction from CDs; *Actual Antenna Accuracy* measures counts from drone imagery on 39 sites.

Runs	Planned Ant. Accuracy	Actual Ant. Accuracy
RUN01	0.615	0.115
RUN02	0.769	0.185
RUN03	0.850	0.231
RUN04	**0.850**	**0.808**

Table 2: Planned accuracy stratified by OCR and whether the antenna table is analyzed. Using both OCR and table analysis yields the best planned-count accuracy.

OCR	Table Analyzed	Planned Acc.	Items
No	No	0.583	39
No	Yes	0.692	39
Yes	No	0.677	39
Yes	Yes	**0.850**	39

after adding OCR, and actual-count accuracy improves substantially once actual counts are derived from a calibrated detector.

Adding OCR increases planned accuracy from 61.5% to 76.9% (+15.4 points). Performance is strongly associated with whether an antenna information table is present on the same page. When extracting from both the table and the antenna plan diagram (with OCR), planned accuracy reaches 85.0% (Table 2).

Using GPT-5 to count antennas directly from drone images is high-variance: across 10 repeated runs on a single site, 9 runs succeeded, and among successful runs the correct actual count was produced in 56% (5/9). Replacing this step with a trained detector reduces variance; after confidence-threshold calibration, end-to-end actual count accuracy reaches 80.8% on the 39-site ANOM1 evaluation set (RUN04).

5.3 Azimuth-Guided Image Retrieval Analysis

Accurate tower GPS is important for selecting images that face a target antenna sector. We measure retrieval quality when tower coordinates are extracted by the agent from construction drawings.

With agent-predicted coordinates, azimuth-guided retrieval attains a mean location error of 7.82 ft and a median error of 5.64 ft between the selected image and the ground-truth reference view. Appendix B provides a detailed ablation over tower-center estimation methods.

5.4 Evaluation of ANOM2 Antenna Dimension Discrepancy Detection

On a limited ANOM2 validation set (two sites; five antennas total), we treat results as *exploratory*. On this set, the pipeline yields low aspect-ratio error (5.3–12.5%) but substantially higher relative error in absolute dimensions (width error 28.7—41.4%, height error 37.4—38.4%). For SITE0001 (2 antennas), width error is 28.7%, height error is 37.4%, and aspect-ratio error is 12.5%. For SITE0002 (3 antennas), width error is 41.4%, height error is 38.4%, and aspect-ratio error is 5.3%. Overall, the results indicate the OBB fitting step is geometrically stable, while monocular metric depth introduces scale uncertainty that dominates absolute dimension error and motivates domain-specific calibration.

Visual analysis of intermediate outputs. Figure 5 shows representative intermediate outputs from ANOM2 (detections, segmentation masks, raw metric depth maps, oriented boxes, and 3D box visualizations) for example tower images.

Across most inspected images, OBB fitting appears stable, suggesting it is usually not the primary failure mode. When it fails, the cause is typically upstream segmentation quality: because OBB fitting depends on SAM [6] masks, partial masks or spillover into background structures can yield unstable contours and degrade the fitted OBB. Future work will evaluate OBB fitting independently of segmentation.

Depth variance is the primary error driver. For SITE0001, two images of the same antenna produce width estimates of 0.88 ft and 1.97 ft (a 2.2× difference), coinciding with large variation in estimated range (10.35 ft vs. 29.67 ft). This sensitivity motivates adaptive depth calibration and multi-view fusion (including newer depth foundation models) to reduce scale variance. In our current experiments we use a fixed depth-range configuration (e.g., maximum depth 40 m), which can be suboptimal when drone-to-tower standoff varies across sites. We expect adapting depth-range parameters using estimated range and capture metadata will reduce saturation/scale artifacts and improve absolute dimension accuracy.

Aspect ratio is comparatively robust. Even when absolute width/height estimates drift, the aspect ratio often remains stable because both dimensions are influenced by a similar depth scale factor. This suggests that aspect ratio can be a more dependable anomaly signal for ANOM2 than absolute dimensions when using monocular depth.

6 CONCLUSIONS

We presented an evidence-grounded, tool-using multimodal agent for plan-to-field anomaly verification in telecom tower audits. The system couples construction drawing understanding with field-image analysis (azimuth-guided retrieval, detection/segmentation, and depth-based 3D measurement) using a rule-driven Detector and a report-only Reviewer. On a 39-site ANOM1 evaluation, the strongest configuration achieves 85.0% planned count accuracy and 80.8% actual count accuracy (RUN04). Using tower coordinates extracted from CDs, azimuth-guided retrieval attains 7.82 ft mean and 5.64 ft median location error. For ANOM2, intermediate outputs (Figure 5) suggest OBB fitting is generally stable, while monocular depth uncertainty dominates absolute-scale error.

Limitations include error propagation from upstream extraction and sensitivity of monocular depth for metric measurements. Future work will improve robustness, incorporate depth calibration and multi-view fusion, and broaden evaluation to include cost/latency and failure-mode metrics.

Beyond accuracy, practical audit workflows benefit from traceability and clear scoping. Our report-only policy and evidence-citation format are intended to support human decision-making rather than automated remediation. To aid reproducibility and iterative development, we preserve per-run execution traces (tool calls, timing, and retry events) alongside the generated report. Because field imagery and site documentation can contain sensitive information, deployments should follow organizational controls for access, retention, and audit logging; our system design is compatible with such controls because it operates on explicitly enumerated evidence artifacts and records what was accessed.

REFERENCES

[1] Christoph Boje, Arianna Guerriero, Sylvain Kubicki, and Yacine Rezgui. 2020. Towards a semantic Construction Digital Twin: Directions for future research. *Automation in Construction* 114 (2020), 103179.

[2] M. Deidda, A. Pala, and G. Sanna. 2020. Modelling a Cell Tower Using SfM: Automated Detection of Structural Elements from Skeleton Extraction on a Point Cloud. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (Archives).*

[3] David H. Douglas and Thomas K. Peucker. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *The Canadian Cartographer* 10, 2 (1973), 112–122.

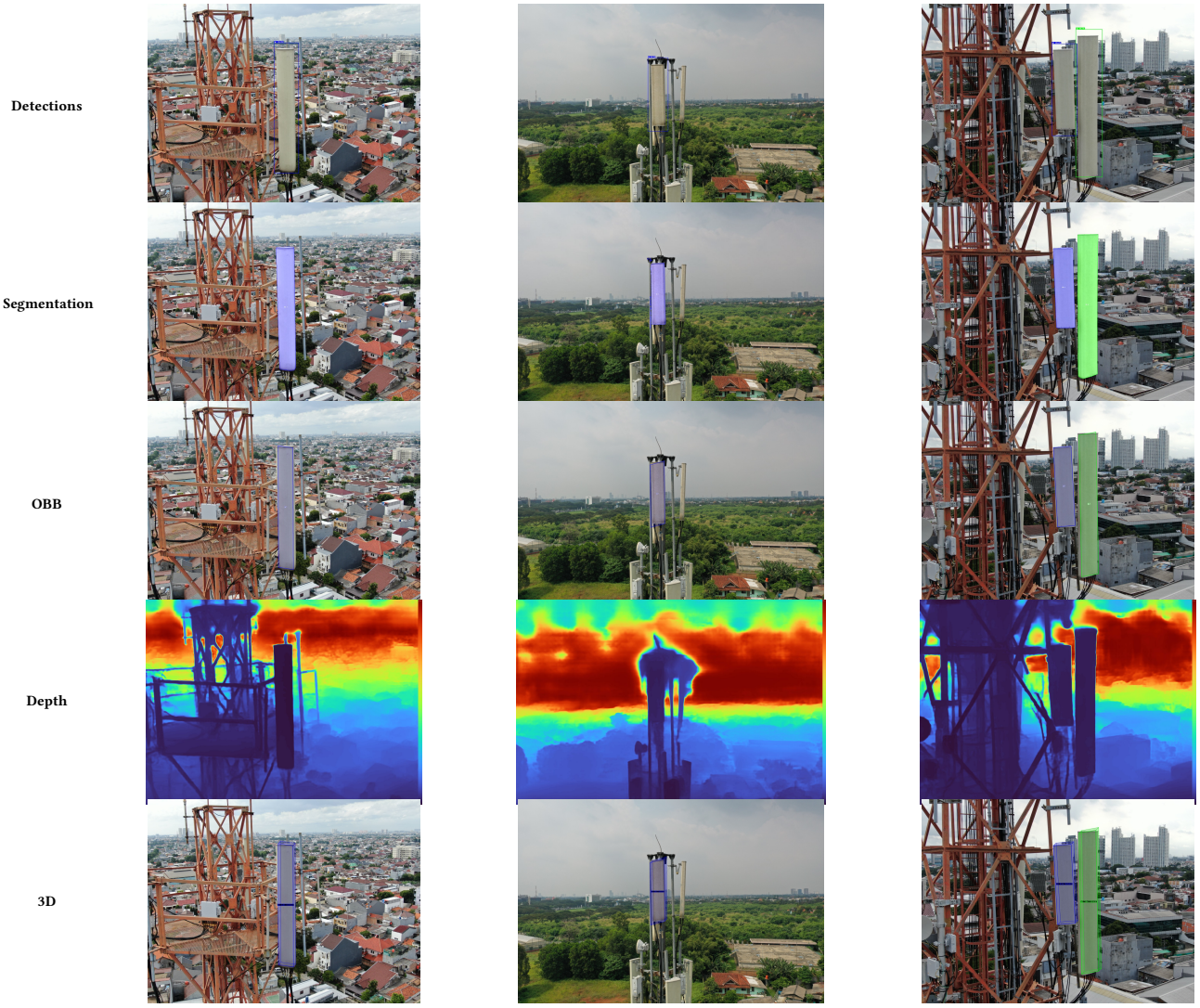


Figure 5: ANOM2 intermediate outputs for three example images. Columns correspond (left to right) to the three example images, and rows show detections, segmentation, oriented boxes, raw metric depth, and 3D visualization.

[4] Z. Hu et al. 2023. AVIS: Autonomous Visual Information Seeking with Large Language Model Agents. In *Advances in Neural Information Processing Systems*.

[5] Jackson Township, NJ. 2015. Cell Tower Plan Sheets Verizon 1CU2015. Online PDF. https://www.jacksontwp.org/wp-content/uploads/2013/03/Cell-Tower-Plan-Sheets-Verizon-1CU2015.pdf Accessed: 2026-01-22.

[6] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 3992–4003.

[7] Werner Kritzinger, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihn. 2018. Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine* 51, 11 (2018), 1016–1022.

[8] P. Lu et al. 2023. Chameleon: Plug-and-Play Compositional Reasoning with Large Language Models. In *Advances in Neural Information Processing Systems*.

[9] Z. Lu, H. Gong, Q. Jin, Q. Hu, and S. Wang. 2022. A Transmission Tower Tilt State Assessment Approach Based on Dense Point Cloud from UAV-Based LiDAR. *Remote Sensing* (2022).

[10] Nicolai Høirup Nielsen. 2025. *YOLO-3D: 3D bounding box estimation with YOLO models*. https://github.com/niconielsen32/YOLO-3D

[11] A. Omairi, Z. H. Ismail, and G. Goycochea Casas. 2025. DF-3DNet: A Lightweight Approach Based on Deep Learning for 3D Telecommunication Tower Asset Classification. *IET Image Processing* (2025).

[12] S. Pu, L. Xie, M. Ji, Y. Zhao, W. Liu, L. Wang, F. Yang, and D. Qiu. 2019. Real-Time Powerline Corridor Inspection by Edge Computing of UAV LiDAR Data. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (Archives)*.

[13] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems*.

[14] Roboflow. 2024. telecom Dataset. https://universe.roboflow.com/project-umuvo/telecom-attennadetection visited on 2026-01-22.

[15] Roger W. Sinnott. 1984. Virtues of the Haversine. *Sky and Telescope* 68, 2 (1984), 159.

[16] Lihe Yang et al. 2024. Depth Anything V2. *arXiv preprint* (2024).

[17] S. Yao et al. 2023. ReAct: Synergizing Reasoning and Acting in Language Models. In *International Conference on Learning Representations (ICLR)*.

[18] Ji Zhang and Sanjiv Singh. 2014. LOAM: Lidar Odometry and Mapping in Real-time. In *Robotics: Science and Systems (RSS)*.

[19] R. Zhang, B. Yang, W. Xiao, F. Liang, Y. Liu, and Z. Wang. 2019. Automatic Extraction of High-Voltage Power Transmission Objects from UAV Lidar Point Clouds. *Remote Sensing* (2019).

APPENDIX

A ANTENNA DETECTOR SCORE THRESHOLD CALIBRATION

We calibrate the Faster R-CNN antenna detector confidence threshold by sweeping a score cutoff and measuring precision/recall trade-offs on a labeled evaluation set (Table 3).

A.1 Object Detection Threshold Calibration

Table 3: Score-threshold calibration results for Faster R-CNN antenna detection models. *Score* is the confidence threshold applied to predicted boxes.

Model	Score	GT	TP	FP	FN	Precision	Recall	F1 Score
v1	0.0	151	150	111	1	0.5747	0.9934	0.7282
v1	0.3	151	150	111	1	0.5747	0.9934	0.7282
v1	0.5	151	150	87	1	0.6329	0.9934	0.7732
v1	0.7	151	150	65	1	0.6977	0.9934	0.8197
v1	0.8	151	150	57	1	0.7246	0.9934	0.8380
v1	0.9	151	149	39	2	0.7926	0.9868	0.8791
v2	0.0	151	149	24	2	0.8613	0.9868	0.9198
v2	0.3	151	149	24	2	0.8613	0.9868	0.9198
v2	0.5	151	149	24	2	0.8613	0.9868	0.9198
v2	0.7	151	149	24	2	0.8613	0.9868	0.9198
v2	0.8	151	149	24	2	0.8613	0.9868	0.9198
v2	0.9	151	149	16	2	0.9030	0.9868	0.9430

B AZIMUTH-GUIDED IMAGE RETRIEVAL: DETAILED ANALYSIS

We analyze azimuth-guided view selection quality for the ANOM1 evaluation pipeline (run `ANOM1_v04_OCR_OBJECT_DETECTION`). The goal is to retrieve, for each target sector azimuth, the drone image that most closely matches a human-selected reference view for antenna counting. We report geodesic distance between retrieved and ground-truth image GPS coordinates (feet), sequence index difference (index), and exact match rate.

Table 4: Azimuth-guided view selection quality on the ANOM1 image retrieval analysis. Lower is better for distance and index difference; higher is better for exact match.

Method	Mean ft	Med. ft	Mean idx	Med. idx	Exact %
Agent predicted	7.82	5.64	5.97	2	10.61
Centroid (GT az.)	5.96	3.98	4.42	2	15.28
Centroid (pred az.)	6.24	3.98	4.64	2	15.71
Hull center (GT az.)	**4.82**	**3.09**	**3.31**	**1**	**22.22**

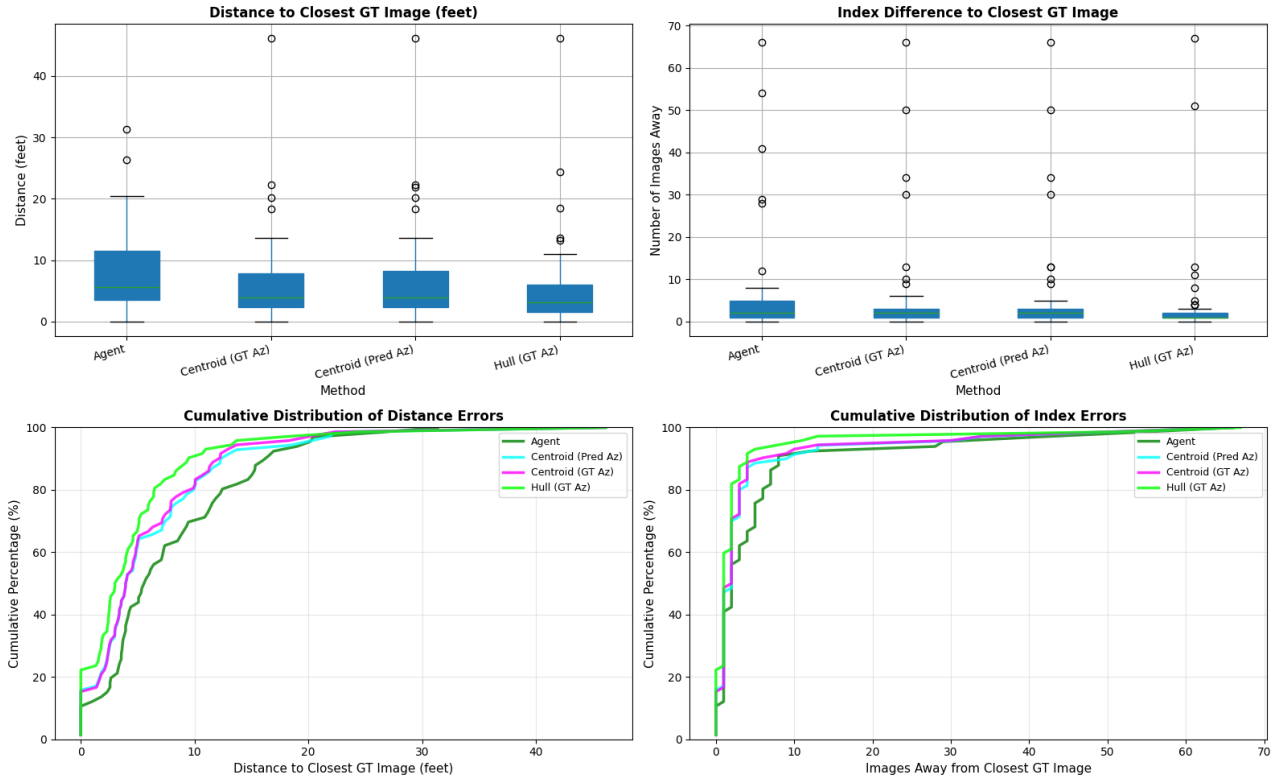


Figure 6: Distribution of retrieval location errors (feet) across methods; convex-hull center yields the tightest error distribution.

C EXAMPLE AGENT-GENERATED REPORT

The following report is a representative example of the structured, evidence-grounded output produced by our agent workflow (formatted in Markdown for readability).

```
# Anomaly Detection Report: SITE001

## Summary
Anomaly analysis result generated by agent workflow.

## Scope & Context
- Site: SITE001
- Date: 2026-01-06T21:04:47.932656+00:00
- Anomaly: ANOM1

## Result
# Anomaly Detection Report: SITE001

## Summary
Using the ANOM1 rule (antenna count discrepancy), I compared planned antenna counts from the FINAL antenna plan diagram in the construction drawing

## Scope & Context
- Site: SITE001
- Date: 2026-01-06
- Context: ANOM1 — Antenna count discrepancy between FINAL CD plan and post-construction imagery.

## Standards & Criteria
- [RULE-ANOM1] Antenna Count Discrepancy: Compare planned counts from FINAL antenna plan diagram(s) vs actual counts from post-construction imager

## Evidence Catalog
- Rule
```

  - [RULE-ANOM1]
- Construction Drawings
  - (CD-PAGE-1): Title Page: Sites/SITE001/cd_page_images/SITE001-FCD_page_1.png
  - (CD-PAGE-6): Elevation Plan: Sites/SITE001/cd_page_images/SITE001-FCD_page_6.png
  - (CD-PAGE-7): Antenna Plan Diagram (FINAL): Sites/SITE001/cd_page_images/SITE001-FCD_page_7.png
- Drone Images (Post-Construction)
  - (IMG-DJI_202508_0214_V): Sites/SITE001/images/postconstruction/Center In/DJI_202508_0214_V.jpeg
  - (IMG-DJI_202508_0230_V): Sites/SITE001/images/postconstruction/Center In/DJI_202508_0230_V.jpeg
- Detection Outputs
  - (DET-0214): detect_antennas on (IMG-DJI_202508_0214_V) → 2 detections; annotated: Sites/SITE001/detection_cache/antenna-anomaly-detection-m
  - (DET-0230): detect_antennas on (IMG-DJI_202508_0230_V) → 4 detections; annotated: Sites/SITE001/detection_cache/antenna-anomaly-detection-m

Analysis
1) Rule initialization and site/CD resolution
- Loaded [RULE-ANOM1] successfully.
- Site info resolved for SITE001; CD at Sites\SITE001\SITE001-FCD.pdf; post-construction imagery available in Center In folder.

2) CD indexing and coordinate extraction
- Index confirms key sheets:
  - (CD-PAGE-1) Title, (CD-PAGE-6) Elevation, (CD-PAGE-7) Antenna Plan Diagram (FINAL), (CD-PAGE-8) Antenna Plan Table, (CD-PAGE-18) Antenna Plan T
- From (CD-PAGE-1), coordinates extracted: latitude 32.7939, longitude -96.945 (validated within ranges). Saved as reference_lat/reference_lon fo

3) Tower elevation (ACL levels)
- From (CD-PAGE-6):
  - ACL1 @ 28 ft (no update), ACL2 @ 72 ft (no update), ACL3 @ 83 ft (proposed_update=true), ACL4 @ 85 ft (proposed_update=true).
  - Proposed updates on ACL3 and ACL4 indicate FINAL antenna diagrams should correspond to these ACLs.

4) FINAL antenna plan diagram(s)
- From (CD-PAGE-7) extraction of FINAL diagram sectors and counts:
  - Returned consistently: Alpha 0°, Beta 120°, Gamma 240°, planned_count=3 each.
  - ACL labeling on the FINAL diagram could not be conclusively read; one pass labeled it as “ACL1” and another as “ACL?” in OCR/vision output. Eleva
  - Proceeded with per-sector planned counts = 3 (Alpha/Beta/Gamma) for the FINAL plan found on (CD-PAGE-7).

5) Post-construction image selection by sector azimuth
- Using reference_lat=42.7939, reference_lon=-86.945, folder: Sites/SITE001/images/postconstruction/Center In
  - Alpha (0° target): best available was (IMG-DJI_202508_0214_V) with view azimuth ~= 133.53° (tolerance ~133.53°)
  - Beta (120° target): (IMG-DJI_202508_0214_V) with view azimuth ~= 133.53° (tolerance ~13.53°)
  - Gamma (240° target): (IMG-DJI_202508_0230_V) with view azimuth ~= 207.35° (tolerance ~32.65°)
- Note: The Alpha sector view is significantly misaligned (by ~133°), so occlusion/visibility is poor for Alpha. Beta is moderately aligned; Gamma i

6) Antenna detections (actual counts per selected image)
- Beta sector (target 120°): (DET-0214) → 2 detected panel antenna faces in (IMG-DJI_202508_0214_V).
- Gamma sector (target 240°): (DET-0230) → 4 detected panel antenna faces in (IMG-DJI_202508_0230_V).
- Alpha sector (target 0°): Only (IMG-DJI_202508_0214_V) available, which is ~=133.53°; detection on this image returns 2 panel faces, but due to ex

7) Comparison (planned vs actual)
- Planned (per FINAL diagram on CD-PAGE-7): Alpha=3, Beta=3, Gamma=3.
- Actual (detections):
  - Beta: actual=2 vs planned=3 → mismatch (diff=1)
  - Gamma: actual=4 vs planned=3 → mismatch (diff=1)
  - Alpha: Indeterminate due to misaligned view (Evidence Gap). A 2-count was observed in the available image, but it cannot be attributed with confi

8) Severity per ANOM1
- Major: Absolute difference = 1 on Beta.
- Major: Absolute difference = 1 on Gamma.
- Minor: Evidence Gap for Alpha sector (indeterminate count due to azimuth misalignment/occlusion).
- Note: Elevation indicates updates at ACL3 and ACL4; however, the FINAL diagram’s ACL labeling is ambiguous in OCR output, preventing per-ACL separ

Findings
ID	Type	Severity	Evidence	Rationale
A1	Compliance Violation	Major	(IMG-DJI_202508_0214_V), (DET-0214), (CD-PAGE-7)	Beta sector: planned 3 vs actual 2 (diff=1).

```
| A2  | Compliance Violation | Major  | (IMG-DJI_202508_0230_V), (DET-0230), (CD-PAGE-7)           | Gamma sector: planned 3 vs actual 4 (diff=1).
| G1  | Evidence Gap      | Minor  | (IMG-DJI_202508_0214_V), (CD-PAGE-7)                     | Alpha sector: no sufficiently aligned post-constru
| G2  | Evidence Gap      | Minor  | (CD-PAGE-6), (CD-PAGE-7)                                 | Elevation shows proposed updates at ACL3 & ACL4, b

## Report Summary (No Notification)
- Generated At: 2026-01-06T00:00:00Z
- Report ID: run-SITE001-ANOM1-7b6c1cde
- Report Path: n/a (analysis session)
- Tools/Models:
  - PDF Indexer/Extractor: CD drawing index tool
  - Vision OCR/Analysis: Azure OpenAI GPT-5 Vision
  - Antenna Detector: antenna-anomaly-detection-model v15_0.00
  - Azimuth-based Image Selector: post-construction image query (Center In)
- Dispatch: none (no notifications sent)

## References
- [RULE-ANOM1]: ANOMALY RULE: ANOM1 — Antenna Count Discrepancy
- (CD-PAGE-1): SITE001-FCD.pdf (page 1) — Sites/SITE001/cd_page_images/SITE001-FCD_page_1.png
- (CD-PAGE-6): SITE001-FCD.pdf (page 6) — Sites/SITE001/cd_page_images/SITE001-FCD_page_6.png
- (CD-PAGE-7): SITE001-FCD.pdf (page 7) — Sites/SITE001/cd_page_images/SITE001-FCD_page_7.png
- (IMG-DJI_202508_0214_V): Sites/SITE001/images/postconstruction/Center In/DJI_202508_0214_V.jpeg
- (IMG-DJI_202508_0230_V): Sites/SITE001/images/postconstruction/Center In/DJI_202508_0230_V.jpeg

## Overall Anomaly Detection Result
ANOMALY_DETECTED=true

## Evaluation
PLANNED_ANTENNA_COUNT=9

ACTUAL_ANTENNA_COUNT=8
```

D ANOMALY RULE DEFINITIONS AND WORKFLOWS USED BY DETECTOR AGENT

We operationalize anomalies as *rule documents*: Markdown specifications that define (i) the rule intent and severity criteria, (ii) the required evidence sources and citations, and (iii) the exact step-by-step tool-driven workflow the Detector must execute (including abort conditions and evidence-gap handling). The MCP server exposes a tool to retrieve these rule documents at runtime; we include two representative rules below verbatim for reproducibility.

D.1 ANOM1 (Antenna Count Discrepancy)

```
# ANOMALY RULE: ANOM1 — ANTENNA COUNT DISCREPANCY

```text

## RULE DESCRIPTION

Use available tools to detect discrepancies between **planned** antenna counts recorded in the **final** Construction Drawings (CDs) and **actual** antenna counts
↪  visible in **post-construction** drone imagery.

## AVAILABLE MCP TOOLS FOR ANALYSIS

1. **get_anomaly_rule_information(anomaly_id)** -- Load rule criteria/plan; must exist for the run to proceed. *(Abort on missing rule.)*
1. **get_site_information(site_id)** -- Resolve `construction_design_path` and image folder map for post construction phases. *(Abort on missing site.)*
1. **get_construction_design_pdf_drawing_index(pdf_path)** -- Extracts drawing index from construction PDFs to identify relevant pages.
1. **get_tower_view_image_by_azimuth(...)** -- Select best post-construction views aligned to sector azimuths to  minimize occlusion.
1. **detect_antennas** - Detect antennas in tower imagery and return bounding boxes with confidence scores.
1. **analyze_image_with_prompt(image_path, prompt, site_id, ocr_enabled)** --  Analyzes an image with a given prompt. Must use valid image paths; produce per-image
↪  counts and visibility/occlusion notes. `ocr_enabled=True` enables OCR on image.

**Evidence citation conventions** -- Use stable identifiers for all evidence:

- **Design sheet:** `[CD-PAGE-<page_number>]`
- **Image:** `[IMG-<basename>]`
- **Rule criteria:** `[RULE-<ID>]`

---

## STEP-BY-STEP DETECTION PROCESS

1. Create `run_id` and validate `(site_id, anomaly_id=ANOM1)`.
1. Use `get_anomaly_rule_information(ANOM1)` to load rule criteria and plan.
  - If NOT_FOUND/HARD_FAIL: produce minimal report and **ABORT** downstream retrieval.
1. Use `get_site_information(site_id)` to obtain `construction_design_path`, image folder map and tower gps coordinates.
  - **ABORT** on NOT_FOUND/HARD_FAIL.
1. Use `get_construction_design_pdf_drawing_index(pdf_path)` to know what the contents of each page are. The filenames for these pages will be used in the image analysis
↪  steps and for citations. If index is missing, continue but note an Evidence Gap. The most important pages are the following:
```
```

  - Final antenna plan diagram
  - Equipment schedule pages
  - Tower elevation plan
  - Title sheet for GPS coordinates (if not returned from `get_site_information`)
1. If and only if the tower gps coordinates is not returned from `get_site_information`, run `analyze_image_with_prompt(image_path, prompt, site_id, ocr_enabled=True)`
↪ on **Title** page image with prompt:

  ```text
  Extract the tower location coordinates (latitude, longitude) from this Title/Site Information page.
  Return decimal degrees. If DMS or degree-minute format is present, convert to decimal.

  Return the result in this exact JSON format:

  {
    latitude: <float>,
    longitude: <float>,
    source_page: <CD-PAGE-x>,
    notes: <any assumptions>
  }

  Example:
  {
    latitude: 40.4461,
    longitude: -79.9822,
    source_page: CD-PAGE-1,
    notes: "Converted from DMS to decimal degrees."
  }
  ```

  - Validate numeric ranges (lat in [-90, 90], lon in [-180, 180]); save as `reference_lat` and `reference_lon`
    for azimuth-guided image selection.

1. On **Tower Elevation** page image run `analyze_image_with_prompt(image_path, prompt, site_id, ocr_enabled=True)` with prompt:

  ```text
  From the tower elevation diagram: The Antenna Center Line level is the ring of equipment like antennas and
  mounts at specific height. List the Antenna Center Line levels in vertical order with acl_level (1 = lowest).
  Denote which ACL level has proposed antenna updates. Updates usually happen to a single ACL level.
  If present, capture ACL/RAD heights.

  Return the result in this exact JSON format:
  {
    "antenna_levels": [
      { "acl_level": int, "rad_height_ft": int, "proposed_update": bool }
    ]
  }

  Example:
  {
    "antenna_levels": [
      { "acl_level": 1, "rad_height_ft": 100, "proposed_update": true },
      { "acl_level": 2, "rad_height_ft": 150, "proposed_update": false },
      { "acl_level": 3, "rad_height_ft": 200, "proposed_update": false }
    ]
  }

  ```

  - Antenna diagrams will only exist for ACL's with `proposed_update: true`.
1. On **Antenna Plan (FINAL)** page image run `analyze_image_with_prompt(image_path, prompt, site_id, ocr_enabled=True)` with prompt:

  ```text
  The following page contains the FINAL antenna plan diagram. It may also contain a tower elevation diagram,
  an EXISTING antenna plan diagram and tables regarding the antenna plans. We will ONLY consider the FINAL
  antenna plan diagram. We want to extract the sector, azimuth, and antenna counts from that diagram only.

  In the FINAL antenna plan diagram:
    - For each sector (Alpha/Beta/...) extract azimuth (degrees) and the count of rectangular panel antennas present.
    - If there are multiple FINAL antenna diagram because updates are on multiple ACL levels (rare), do the extraction per ACL level. There should only be diagrams for
    ↪  the ACL being updated: <insert ACL levels where proposed_update=true>.

  Return a JSON object structured as:
  {
    "<acl_level>": {
      "<sector_name>": {
          "azimuth_degrees": <number>,
          "planned_count": <number>
      }
    }
  }

  Example:
  {
    "ACL1": {
      "Alpha": { "azimuth_degrees": 0, "planned_count": 2 },
      "Beta": { "azimuth_degrees": 120, "planned_count": 2 },
      "Gamma": { "azimuth_degrees": 240, "planned_count": 2 }
    },
    "ACL2": {
      "Alpha": { "azimuth_degrees": 130, "planned_count": 1 }
    }
  ```

  }
  ```

  This enforces **final-only** and builds the planned per-ACL, per-sector counts.

1. For each sector, call `get_tower_view_image_by_azimuth(...)` with:

  - `reference_lat`, `reference_lon`,
  - `target_azimuth` = sector's azimuth (degrees),
  - `site_id`
  - `construction_phase="postconstruction"`,
  - `folder_name` using the path in `drone_image_folders` returned from `get_site_information`. Eg: `Sites/<site_id>/images/postconstruction/Center In/`

  This will select the **best** post-construction image aligned with the target azimuth in order to minimize occlusion for that sector.
  Collect 1 view per sector/ACL.

1. For each image returned from the previous step, use `detect_antennas(image_path, site_id, azimuth, score_threshold)` to detect the antennas present in the images. The number of detected antennas will determine the number of antennas counted in the image.

  - `image_path`: image filepaths returned from previous step
  - `site_id`
  - `azimuth`: azimuths returned from previous step
  - `score_threshold`: use 0.88 as the threshold

In the response, the `metadata` field will contain `number_of_detections`.
This is "actual" antenna count for that azimuth sector which we will to the planned antenna counts from earlier.

1. For each for each results above, compare `planned_count` (from FINAL plan diagram) vs `actual_count` (drone).
  Record **mismatch** findings and **Evidence Gaps** where counting couldn't be completed with confidence
  (occlusion/quality).

1. Severity (rule-specific for ANOM1):
  - **Critical** -- Absolute difference in antenna count between planned and actual antenna count >= 2 on any
    ACL, or mismatches across multiple sectors on the same ACL.
  - **Major** -- Absolute difference in antenna count between planned and actual antenna count = 1 on any ACL.
  - **Minor** -- Required count indeterminate due to evidence quality -> **Evidence Gap** recorded (not a pass).

## Reporting (Report-Only)

1. **Evidence Catalog (with citations)** -- list every CD page `(CD-PAGE-x)`, every image `(IMG-...)`, and any
  analysis sets with stable IDs.
1. **Analysis narrative** -- reference your citations while explaining how counts were derived and reconciled.
1. **Findings table** -- ID | Type | Severity | Evidence | Rationale (no remediation text).
1. **Report Summary (No Notification)** -- Timestamp, run ID/path, tool versions; explicitly state **no
  notifications sent**.
1. **References** -- map `(IMG-*)` and `(CD-Page-*)` to file names/paths/pages.

> The Reviewer checks that each mandatory step was performed and that the "no remediation" policy is respected.
> If anything's missing, they return **MISSING_STEPS** with targeted feedback; otherwise **COMPLETE**.

## Failure & Gap Handling (Detector policy)

- **NOT_FOUND/HARD_FAIL** on rule/site -> minimal report + **ABORT** downstream.
- **Model errors** on visual analysis -> 1 retry; if still failing, record **Evidence Gap** and continue (never
  hallucinate).
- Every HARD_FAIL maps to **Missing Context** (essential) or **Evidence Gap** (other) so absence is explicit.

---

### Markdown Report Format (for the report field)

```` text
# Anomaly Detection Report -- <Site ID>

## Summary
<one-paragraph summary>

## Scope & Context
- Site: <ID>
- Date: <YYYY-MM-DD>
- Context: <brief>

## Standards & Criteria
- <Standard/Rule 1>
- <Standard/Rule 2>

## Evidence Catalog
- (IMG-12): Drone image DJI_0392.JPG
- (CD-Page-7): Construction Design PDF page 7

## Analysis
Discuss observations referencing (IMG-12), (CD-Page-7), etc.

## Findings
| ID | Type                  | Severity | Evidence               | Rationale |
|----|----------------------|----------|------------------------|----------|
| A1 | Compliance Violation | Major    | (IMG-12), (CD-Page-7) | <why>     |

## Report Summary (No Notification)
- Generated At: <UTC timestamp>
- Report ID: <ID or hash>

- Report Path: <path or URL>
- Dispatch: none (no notifications sent)

## References
(IMG-12): DJI_0392.JPG (image)
(CD-Page-7): CDs.pdf (page 7)

## Overall Anomaly Detection Result
ANOMALY_DETECTED=false if the total planned antenna counts (PLANNED_ANTENNA_COUNT) matches the total actual antenna count (ACTUAL_ANTENNA_COUNT), otherwise true; in case of an evidence gap, always set this to true.

## Evaluation
PLANNED_ANTENNA_COUNT=<Total planned antenna counts as integer only>

ACTUAL_ANTENNA_COUNT=<Total actual antenna counts as integer only>
```

D.2 ANOM2 (Antenna Physical Design Verification)

```
# ANOMALY RULE: ANOM2 -- ANTENNA PHYSICAL DESIGN VERIFICATION

## RULE DESCRIPTION

Verify that **actual** antenna physical dimensions in **post-construction** drone imagery match the **planned** specifications from **final** Construction Drawings (CDs). This rule uses 3D bounding box detection to extract precise measurements of antenna dimensions (height, width, depth) and distance. Comparisons are made at the **ACL** level and **sector** level when available.

## MEASUREMENT CONVENTIONS (ANOM2)

- **Planned specifications** -- From **final** CD antenna plan diagram/table and equipment schedules; cite page numbers.
- **Actual measurements** -- Extracted from post-construction drone images using 3D bounding box detection.
- **ACL (Antenna Center Line)** -- Vertical center height of an antenna (often labeled **RAD height**) in the CDs.
- **ACL level** -- A vertical tier of antennas at the same elevation on the structure (1 = lowest).
- **Sector & azimuth** -- Sector labels (Alpha/Beta/Gamma/Delta) and `azimuth_degrees` from antenna plans.
- **Dimensions** -- Physical antenna dimensions (height, width, depth) in feet.
- **Distance** -- Distance from camera to antenna in feet, measured using monocular depth estimation.

## AVAILABLE MCP TOOLS FOR ANALYSIS

1. **get_anomaly_rule_information(anomaly_id)** -- Load rule criteria/plan; must exist for the run to proceed.
  *(Abort on missing rule.)*
1. **get_site_information(site_id)** -- Resolve `construction_design_path` and image folder map for post construction phases. *(Abort on missing site.)*
1. **get_construction_design_pdf_drawing_index(pdf_path)** -- Extracts drawing index from construction PDFs to identify relevant pages.
1. **get_tower_view_image_by_azimuth(...)** -- Select best post-construction views aligned to sector azimuths to minimize occlusion.
1. **generate_3d_bounding_boxes(image_path, site_id)** -- Detects antennas and enriches with oriented bounding boxes (OBB), distance estimation, dimensions, and 3D visualization.
1. **analyze_image_with_prompt(image_path, prompt)** -- Use for extracting planned specifications from CD pages.

**Evidence citation conventions** -- Use stable identifiers for all evidence:

- **Design sheet:** `[CD-PAGE-<page_number>]`
- **Image:** `[IMG-<basename>]`
- **Detection:** `[DETECTION-<antenna_id>]` (use for both basic detections and 3D analysis)
- **Rule criteria:** `[RULE-<ID>]`

---

## STEP-BY-STEP DETECTION PROCESS

## PHASE 1 -- Scope & Initialize

1. Create `run_id` and validate `(site_id, anomaly_id=ANOM2)`.
1. `get_anomaly_rule_information(ANOM2)` -> load rule criteria and plan. If NOT_FOUND/HARD_FAIL: produce minimal report and **ABORT** downstream retrieval.

## PHASE 2 — Load Site Context & Index/Rasterize CDs

1. Use `get_site_information(site_id)` to obtain `construction_design_path` and image folder map. **ABORT** on NOT_FOUND/HARD_FAIL.
1. Use `get_construction_design_pdf_drawing_index(pdf_path)` to identify the contents of each page. The most important pages are:
  - **Final** antenna plan diagram
  - Equipment schedule pages (antenna specifications)
  - Tower elevation plan
  - Title sheet for GPS coordinates

The filenames for these pages will be used in the image analysis steps and for citations. If index is missing, continue but note an **Evidence Gap**.

## PHASE 3 — Extract **Planned** Specifications from FINAL CD Pages

1. On **Title** page image run `analyze_image_with_prompt(image_path, prompt)` with prompt:

  ```text
  "Extract the tower location coordinates (latitude, longitude) from this Title/Site Information page.
  Return decimal degrees. If DMS or degree-minute format is present, convert to decimal. Output:
  { latitude: <float>, longitude: <float>, source_page: <CD-PAGE-x>, notes: [any assumptions] }."
  ```

  - Validate numeric ranges (lat in [-90, 90], lon in [-180, 180]); save as `reference_lat` and `reference_lon` for azimuth-guided image selection.

1. On **Antenna Plan (FINAL)** page image run `analyze_image_with_prompt(image_path, prompt)` with prompt:

  ```text
  "From this FINAL antenna plan diagram: for each sector (Alpha/Beta/...), extract azimuth (degrees) and antenna count per ACL level (acl_level, 1 = lowest).
```

  Ignore diagrams with 'existing' or non-final annotations. Return JSON keyed by acl_level and sector:
  { acl_level: { sector: { azimuth_degrees, antenna_count } } }."
  ```

1. On **Equipment Schedule** page image run `analyze_image_with_prompt(image_path, prompt)` with prompt:

  ```text
  "From this equipment schedule: extract antenna model numbers and physical dimensions (height, width, depth in feet).
  Return JSON: [ { model: <string>, height_ft: <float>, width_ft: <float>, depth_ft: <float> }, ... ]."
  ```

1. On **Tower Elevation** page image run `analyze_image_with_prompt()` with prompt:

  ```text
  "From this tower elevation: list ACL levels in vertical order with acl_level (1 = lowest).
  If present, capture ACL/RAD heights. Output:
  [ { acl_level: 1, rad_height_ft: <float or null> }, ... ]."
  ```

PHASE 4 -- Acquire **Post-Construction** Drone Evidence (by sector azimuth)

1. For each sector, call `get_tower_view_image_by_azimuth(...)` with:

  - `reference_lat`, `reference_lon`,
  - `target_azimuth` = sector's azimuth (degrees),
  - `site_id`, `construction_phase="postconstruction"`, and the appropriate `folder_name`.

  `folder_name` will be the full filepath to the image locations. Ex. `Sites/SITE0001/images/postconstruction/Center In`.
  This will select the **best** post-construction image aligned with the target azimuth to minimize occlusion for that sector.
  Collect 1 view per sector/ACL.

PHASE 5 — Detect & Measure **Actual** Antennas in Drone Images

For each selected image:

1. Step 5.1: Detect and Measure Antennas

  Call `generate_3d_bounding_boxes(image_path, site_id)` to detect antennas and compute physical measurements:

  ```json
  {
    "status": "success",
    "detections": [
     {
      "id": 1,
      "bbox": [100, 100, 200, 400],
      "obb": [[100, 100], [200, 100], [200, 400], [100, 400]],
      "confidence": 0.95,
      "dimensions": {
        "height_in_feet": 6.0,
        "width_in_feet": 1.5,
        "depth_in_feet": 0.7
      },
      "distance": {
        "median_distance": 53.1,
        "mean_distance": 53.5,
        "min_distance": 51.8,
        "max_distance": 54.8,
        "std_distance": 0.75
      },
      "class_name": "panel antenna"
     }
    ],
    "visualization_path": "Sites/SITE0001/intermediate-image-results/3d_boxes_DJI_0392.jpg"
  }
  ```

  - Validate detection count matches planned specifications (+/-2 tolerance).
  - Extract `dimensions` for physical size verification.
  - Extract `distance.median_distance` for spatial positioning analysis.
  - Cite as `[DETECTION-{id}]` for evidence.

1. Step 5.2: Map Detections to Planned Specifications

  For each detection:
  - Verify `dimensions` match equipment schedule specifications (within $\pm 10\%$ tolerance).
  - Note any antennas that cannot be confidently matched.

PHASE 6 -- Planned vs. Actual Comparison (ANOM2)

For each **ACL -> sector**, compare:

1. **Dimensional Verification:**
  - Compare actual `dimensions` (height, width, depth) vs. equipment schedule specifications.
  - **Mismatch** if any dimension differs by > 10%.

1. **Count Verification:**
  - Compare `number_of_detections` vs. planned antenna count.
  - **Mismatch** if counts differ.

Severity (rule-specific for ANOM2):
- **Critical** -- Dimensional mismatch > 20% suggesting wrong equipment model installed, or antenna count mismatch.
  ```

- **Major** -- Dimensional mismatch 10-20%.
- **Minor** -- Dimensional mismatch < 10%.
- **Evidence Gap** -- Detection failed, distance estimation unavailable, or image quality insufficient for measurement.

PHASE 7 -- Synthesis & Reporting (Report-Only)

1. **Evidence Catalog (with citations)** -- list every CD page `(CD-PAGE-x)`, every image `(IMG-...)`, and every detection `(DETECTION-x)` with stable IDs.
1. **Analysis narrative** -- reference your citations while explaining how measurements were derived and compared.
1. **Findings table** -- ID | Type | Severity | Evidence | Rationale (no remediation text).
1. **Visualizations** -- Include paths to 3D bounding box visualization images for visual evidence.
1. **Report Summary (No Notification)** -- Timestamp, run ID/path, tool versions; explicitly state **no notifications sent**.
1. **References** -- map `(IMG-*)`, `(CD-Page-*)`, and `(DETECTION-*)` to file names/paths/pages.

> The Reviewer checks that each mandatory step was performed and that the "no remediation" policy is respected.
> If anything's missing, they return **MISSING_STEPS** with targeted feedback; otherwise **COMPLETE**.

Failure & Gap Handling (Detector policy)

- **NOT_FOUND/HARD_FAIL** on rule/site -> minimal report + **ABORT** downstream.
- **Model errors** on detection -> 1 retry; if still failing, record **Evidence Gap** and continue (never hallucinate).
- **Depth estimation failure** -> continue with detection data only, note **Evidence Gap** for dimensional analysis.
- Every HARD_FAIL maps to **Missing Context** (essential) or **Evidence Gap** (other) so absence is explicit.

REFERENCES

- **Ground Truth Curation** -- authoritative **final** CD plans; equipment schedule specifications; ACL/sector/azimuth schema; dimensional tolerances.
- **Agent Design** -- MCP skills; Detector/Reviewer responsibilities; evidence citation conventions; workflow; failure policy; report/JSON contracts.
- **3D Bounding Box Design** -- Detection methodology; distance estimation calibration; visualization outputs.

Report must follow the format below:

```text
# Anomaly Detection Report: <Site ID>

## Summary
<one-paragraph summary>

## Scope & Context
- Site: <ID>
- Date: <YYYY-MM-DD>
- Context: <brief>

## Standards & Criteria
- <Standard/Rule 1>
- <Standard/Rule 2>

## Evidence Catalog
- (IMG-12): Drone image DJI_0392.JPG
- (CD-Page-7): Construction Design PDF page 7
- (DETECTION-1): Antenna detection ID 1 with 3D measurements
- (VIS-12): 3D visualization Sites/SITE0001/intermediate-image-results/3d_boxes_DJI_0392.jpg

## Analysis
Discuss observations referencing (IMG-12), (CD-Page-7), (DETECTION-1), etc.

## Findings
| ID | Type                 | Severity | Evidence                          | Rationale |
|----|----------------------|----------|-----------------------------------|-----------|
| A1 | Dimensional Variance | Major    | (DETECTION-1), (CD-Page-7)        | <why>     |
| A2 | Distance Mismatch    | Minor    | (DETECTION-2), (CD-Page-8)        | <why>     |

## Report Summary (No Notification)
- Generated At: <UTC timestamp>
- Report ID: <ID or hash>
- Report Path: <path or URL>
- Dispatch: none (no notifications sent)

## References
(IMG-12): DJI_0392.JPG (image)
(CD-Page-7): CDs.pdf (page 7)
(DETECTION-1): Antenna ID 1, confidence 0.95, dimensions {height: 6.0, width: 1.5, depth: 0.7} ft, distance 53.1 ft
(VIS-12): Sites/SITE0001/intermediate-image-results/3d_boxes_DJI_0392.jpg (3D visualization)

## Evaluation
PLANNED_ANTENNA_COUNT=<Total planned antenna counts as integer only>

ACTUAL_ANTENNA_COUNT=<Total actual antenna counts from detections as integer only>

DIMENSIONAL_MISMATCHES=<Number of dimensional mismatches as integer only>
```