Tribunal Session Summary

Session: conclave-20260620-135847-for-a-hyperscaler-build-8ba2

Depth: T6

Advocates: 6 | Judges: 5 | Cost: $5.4836 | Time: 26m 53s

THE PROMPT: For a hyperscaler build out, what are the top key build out milestones that would be identifiable from 30cm satellite imagery and support the verification of common publicly reported milestones?

THE RULING: The top key build-out milestones identifiable from 30cm satellite imagery are Substation/Grid-Tie Completion, Generator Installation, Shell Completion, Steel Framing, and Foundation Completion, ranked by verification value. Power infrastructure milestones (substation and generators) provide the highest falsification power for high-stakes public claims like "energized" or "online," while structural milestones verify lower-stakes progress claims. This framework enables "negative verification" - using imagery to definitively falsify premature claims rather than confirm operational status.

The Prompt

For a hyperscaler build out, what are the top key build out milestones that would be identifiable from 30cm satellite imagery and support the verification of common publicly reported milestones?

The Ruling

The top key build-out milestones identifiable from 30cm satellite imagery are Substation/Grid-Tie Completion, Generator Installation, Shell Completion, Steel Framing, and Foundation Completion, ranked by verification value. Power infrastructure milestones (substation and generators) provide the highest falsification power for high-stakes public claims like "energized" or "online," while structural milestones verify lower-stakes progress claims. This framework enables "negative verification" - using imagery to definitively falsify premature claims rather than confirm operational status.

The strongest evidence comes from Aterio's operational methodology, which explicitly tracks these five milestones across U.S. hyperscale campuses, and the Epoch AI/FAS frameworks that treat substation completion as the binding constraint for energization. The main caveat is that imagery can only verify external physical readiness, not internal commissioning or operational status.

Core Analysis

The Ranked Verification Hierarchy

The most effective framework for verifying hyperscale build-outs prioritizes milestones by their ability to falsify premature or inaccurate public disclosures. The primary verification logic is "negative verification" - imagery is more reliable at proving a claim is physically impossible than at proving hidden internal states.

Rank	Milestone	Public Claim Verified	Imagery Signature (30 cm)	Verification Logic & Confidence
1	**Substation / Grid■Tie Completion**	"Campus energized," "Grid connection ready," "First building online"	Fully built high■voltage yard, visible transformers, buswork, and terminated transmission lines	**Constraint Anchor:** Binding campus-side constraint for energization. Physical absence definitively falsifies "energized" claims. (*Confidence: High*)

Rank	Milestone	Public Claim Verified	Imagery Signature (30 cm)	Verification Logic & Confidence
2	**Generator Installation**	"Backup power installed," "Critical infrastructure ready"	Discrete units (e.g., Caterpillar C175■16) on dedicated concrete pads; countable banks	**Directional Proxy:** Count provides falsifiable upper bound for backup capacity. Verifies physical installation, not functional commissioning. (*Confidence: Medium■High*)
3	**Shell Completion**	"Building watertight," "Fabrication complete"	Enclosed roof and walls; absence of open bays or visible structural skeleton	**Structural End State:** Last purely external, capital-intensive structural milestone. Distinguishable from earlier framing stages. (*Confidence: High*)
4	**Steel Framing**	"Vertical construction underway," "Topped out"	Full■height steel skeleton footprint and visible structural height	**Progress Marker:** Verifies vertical progression but cannot confirm operational readiness. Aggregated roofline is the signal, not individual columns. (*Confidence: Medium■High*)
5	**Foundation Completion**	"Ground broken," "Construction started"	Visible excavation, rebar, and poured foundation pads	**Irreversible Start:** Distinguishes permanent construction start from reversible site clearing/grading. (*Confidence: High*)

Verification Mechanics and Falsification Logic

The framework applies imagery as a **corroboration/falsification layer** rather than standalone proof of operation:

- **Necessary, Not Sufficient:** Substation and generators are *necessary* conditions for "energized" status but not *sufficient* (internal fit-out and utility commissioning are invisible)
- **Falsification Power:**
  - High-stakes claims ("Energized"): Require verification of Power Milestones (#1-2)

  - Progress claims ("Under Construction"): Require verification of Structural Milestones (#3-5)
- **Generator Count as Upper Bound:** A count of generators establishes maximum theoretical backup capacity but cannot account for invisible redundancy configurations (N+1/2N)
- **Temporal Confirmation:** Multi-date temporal stacks are required to confirm progression and mitigate single-image false positives

Technical Constraints and Boundary Conditions

- **Resolution Limits:** At 30cm GSD, individual steel columns (0.3-0.6m) are often below reliable detection. Verification relies on aggregate structural features
- **"Readiness" Definition:** Imagery can only verify **physical installation completion**, not functional commissioning
- **Campus Variation:** Hyperscale campuses are phased. Verification requires mapping different buildings to different states (e.g., Building A at Shell, Building B at Foundation)

Evidence Base

The framework draws on three key sources:

1. **Aterio's Operational Methodology:** Explicitly tracks foundations, steel framing, shell completion, generator installation, and substation build-out across U.S. hyperscale campuses
2. **Epoch AI's "Build Times" Framework:** Aligns satellite observation of substation build-out with power milestones toward gigawatt campuses
3. **FAS "Tracking Hyperscale AI Data Center Growth":** Treats substation and grid-tie infrastructure as key verifiable features in imagery

How We Got Here

Scorecard

Judge	Claude Opus 4.8	GPT-5	Gemini 3.1 Pro	DeepSeek V4 Pro	Grok 4	Magistral Medium
DeepCogito v2.1	1	2	5	4	3	6
GPT-OSS 120B	2	1	3	4	5	6
Mistral Large 3	1	2	3	4	5	6
Zhipu GLM-5	2	1	3	4	5	6

Judge	Claude Opus 4.8	GPT-5	Gemini 3.1 Pro	DeepSeek V4 Pro	Grok 4	Magistral Medium
Zhipu GLM-4.7	2	1	3	4	5	6
Winner	**Claude Opus 4.8**	**GPT-5**	**Gemini 3.1 Pro**	**DeepSeek V4 Pro**	**Grok 4**	**Magistral Medium**

Convergence Assessment

The convergence was predominantly epistemic, driven by evidence and logical argument rather than social pressure. Key shifts were evidence-based:

- Abandonment of "just-in-time" capital efficiency premise when contradicted by market reality
- Adoption of "negative verification" framework based on imagery's inherent limitations
- Recognition of substation as the binding constraint after examining dependency chains
- Technical corrections on resolution limits (steel columns, generator distinctness)

Position changes were triggered by verifiable corrections (e.g., Aterio's methodology explicitly distinguishing foundation work from site clearing) rather than social dynamics.

Closing Statements

Claude Opus 4.8 Finalized a ranked framework with substation #1 and generators #2, emphasizing falsification power over high-stakes claims. Conceded BESS visibility challenges but maintained the core power-first verification logic. Strongest on analytical rigor and evidence integration.

GPT-5 Evolved from an overinclusive 11-milestone list to a precise ranked hierarchy, incorporating the group's best insights on falsification mechanics. Provided the cleanest synthesis of verification logic and technical constraints.

Gemini 3.1 Pro Pivoted dramatically from an initial "just-in-time" thesis to embrace negative verification, demonstrating intellectual honesty. The pivot helped crystallize the consensus framework around falsification power.

DeepSeek V4 Pro Provided crucial technical rigor on resolution limits and boundary conditions. Successfully defended the necessity of ranking while conceding the need to split power infrastructure into distinct milestones.

Grok 4 Started with a flat milestone list but eventually adopted the ranked hierarchy. Contributed strong source fidelity to Aterio's methodology but was slower to embrace the falsification framework.

Magistral Medium Maintained a basic milestone list throughout, adding mapping logic under pressure but never developing a substantive verification framework. Failed to engage with the core challenge of how imagery supports verification.

Key Moments

- **Claude Opus 4.8's "Binding Constraint" Argument (Round 3)** Claude Opus 4.8 established that substation completion is the binding campus-side constraint for energization, forcing other advocates to confront the logical priority of power infrastructure over structural milestones. This became the foundation for the ranked hierarchy.
- **Gemini 3.1 Pro's Pivot to "Negative Verification" (Round 2)** Gemini 3.1 Pro abandoned its initial "just-in-time" capital efficiency premise when challenged with evidence of pre-positioning, adopting instead the "negative verification" framework that would become central to the consensus.
- **DeepSeek V4 Pro's Technical Correction on Steel Resolution (Round 5)** DeepSeek V4 Pro conceded that individual steel columns are not reliably classifiable at 30cm, shifting the verification signal to aggregate roofline features and establishing important technical boundaries for the framework.
- **GPT-5's Pruning of Low-Value Milestones (Round 4)** GPT-5 removed site mobilization and security indicators from its initial 11-milestone list after challenges about their reversibility and low verification value, converging on the core five milestones.
- **The BESS Succession Debate (Round 7)** Claude Opus 4.8 conceded that battery energy storage systems lack the distinct visual signature of diesel generators at 30cm, acknowledging uncertainty about future verification anchors as the industry shifts away from diesel.

Reality Checks

Claim	Challenge	Result
Gemini 3.1 Pro: "Just-in-time capital efficiency makes generators a proxy for operational status"	**Claude Opus 4.8:** Market evidence shows hyperscalers pre-position power infrastructure for grid queue position	Full concession; thesis abandoned in Round 2
DeepSeek V4 Pro: "30cm resolves individual steel columns (0.3-0.6m)"	**Claude Opus 4.8:** At 30cm GSD, a 0.3m column is a single pixel - below reliable classification	Conceded; revised to aggregate roofline features
GPT-5: "Site mobilization verifies groundbreaking"	**DeepSeek V4 Pro:** Aterio methodology explicitly distinguishes reversible site clearing from foundation work	Conceded; removed from milestone list
Claude Opus 4.8: "BESS yards will inherit generator-count verification logic"	**Gemini 3.1 Pro:** BESS containers lack distinctive features and may be centrally aggregated, breaking building-level counting	Conceded "clean transfer" claim; confidence downgraded to ~50%

Next Steps

Next Step	Why It Matters	Confidence
Develop standardized discrimination criteria for BESS/fuel-cell yards at 30cm	As diesel backup declines, verification needs successor anchors	60%
Quantify typical lag between substation completion and actual energization	Improves interpretation of "energized" claims	70%
Map off-campus utility constraints that affect energization timing	Identifies key blind spot in current verification	65%
Establish confidence bounds for generator count → MW conversion	Strengthens capacity verification	75%

Dissenting Opinions

No dissenting opinions were filed in this session.

Glossary

Term	Definition
30cm GSD	Ground Sample Distance - each pixel represents 30cm on the ground, standard for high-resolution commercial satellite imagery
Binding Constraint	The limiting factor that determines when a milestone can be achieved; for energization, the substation is the binding constraint
Falsification Power	The ability of a milestone observation to definitively disprove a public claim
Negative Verification	Using imagery to prove a claim is false (e.g., no substation = not energized) rather than confirming it is true
N+1/2N Redundancy	Backup power configurations that affect how generator count translates to usable capacity; invisible in imagery
Phased Campus	A multi-building hyperscale development constructed in stages, with different buildings at different completion states
Temporal Stack	Multiple satellite images of the same location across different dates, used to confirm progression

How The Conclave Works

The Conclave is a structured multi-model deliberation system. Instead of asking one AI model a question and trusting its answer, the Conclave convenes a panel of independent models, forces them to argue, and subjects their conclusions to judicial review. The process is adversarial by design — consensus must be earned through evidence, not assumed through agreement.

Advocates are independent AI models (e.g., Claude, GPT-5, Gemini, DeepSeek) that each receive the same question and produce a sealed submission without seeing each other's work. Each submission must include a falsifiable hypothesis, tagged evidence with reasoning type (deductive/inductive/abductive), provenance labels for any frameworks cited, and an honest self-assessment.

Challenges follow submissions. Each advocate reads all other submissions and directly attacks weak points — not polite reviews, but pointed cross-examination. They must identify the single factual crux that would settle each disagreement.

Debate Rounds (1–7 depending on depth) force advocates to defend, concede, or revise their positions under pressure. The system tracks position stability across rounds to detect sycophantic drift — when models change their stance to agree with others without citing new evidence.

Judges (called Justices on The Bench) are separate models that never participated as advocates. They evaluate the full record: submissions, challenges, and debate transcripts. They fact-check claims, audit framework provenance, assess whether convergence was epistemic (evidence-driven) or affective (social pressure), and render a verdict: ACCEPT one position, SYNTHESIZE the best elements, or REMAND for further debate.

Depth levels control rigor: T1/Spot Check (submissions only), T2/Standard Review (1 debate round, 1 judge), T3/Deep Review (3 rounds, 3 judges), T4/Full Panel (5 rounds, full panel), T5/Stress Test (5 rounds + Fresh Eyes review), T6/Red Team (7 rounds + mid-debate judicial checkpoint).

The system is deterministic code — it cannot be sycophantic. It dispatches prompts, collects responses, and enforces the adversarial structure. The models argue; the code referees.