

Claims

Our Place in the Bittensor Ecosystem

Scientific Data, Verification Mechanisms, and Comparative Design

Bittensor Subnet 111

Claims Team

Philipp Koellinger

Professor of Economics, Vrije University

Christian Roessler

Associate Professor of Economics, California State University-East Bay

Ogban Ugot

Engineer

Comparison Paper v1 • Draft release snapshot • August 2026

Purpose. This paper places CLAIMS in the Bittensor ecosystem by separating two kinds of proximity: the product a subnet creates and the mechanism it uses to verify and reward work. It compares SN111 with related scientific, data, retrieval, and benchmark systems, keeps the active v0 mechanism distinct from proposed extensions, and identifies which established design patterns transfer directly to scientific extraction and which must change when semantic correctness is only partly verifiable.

Abstract

Claims is a Bittensor network that turns scientific papers into persistent, source-linked records of claims and evidence. This paper compares its product and incentive design with nearby scientific, data, retrieval, and verification networks. The closest overall analog is Minos, a genomic benchmarking network that also combines private methods, hidden evaluation, and reusable scientific output. Yet genomic tests can use concealed synthetic answer keys, whereas scientific extraction must also handle omissions, equivalent wording, qualifications, evidence relations, and genuine ambiguity. Claims therefore uses direct source checks, meaning-based adjudication, and several levels of review.

Keywords: scientific data extraction; claim–evidence graphs; Bittensor; decentralized AI; subnet comparison; selective verification; synthetic challenges; contribution markets.

Contents

1	Dimensions of Comparison	1
2	Claims as the Reference Architecture	4
3	Scientific and Technical Comparators	8
4	Data Construction and Semantic Extraction	12
5	Search and Retrieval: Substitute or Complement?	16
6	Truth Structures and Verification	17
7	Competition, Rewards, and Independent Information	21
8	Validators, Integrity, and Decentralization	25
9	Closest Comparisons on a Common Template	28
10	Operating and Market Context	33
11	Conclusion	37
A	Relevant-Subnet Comparison Matrix	38
B	Comparator Profiles	41
C	Claims Current–Prospective Crosswalk	42
D	Detailed Market Table	43
E	Glossary	45

1 Dimensions of Comparison

In markets for machine intelligence, useful output needs to be recognized and rewarded. Scientific curation makes that recognition unusually difficult. A record may be well formed yet omit a central result, attach the wrong evidence, collapse distinct findings, or state more than a paper supports. Some errors can be checked directly against a source; others require judgment, and some remain genuinely unresolved.

Within Bittensor, two types of comparisons are especially plausible. Subnets may resemble each other because both produce neighboring products, or because both solve neighboring problems of verification and incentives. Those relationships often diverge. CLAIMS is a scientific-data system, a structured-extraction system, and a market for machine work whose correctness is only partly observable. Its position becomes clear only when product proximity and mechanism proximity are examined separately.

1.1 Product proximity and mechanism proximity

Product proximity concerns what remains after a task is complete. The relevant distinctions include whether the output is a transient response or a persistent asset; whether it remains linked to its sources; whether it can be corrected; whether downstream systems can reuse it; and whether its value comes mainly from acquisition, labeling, generation, retrieval, or canonical representation.

Mechanism proximity concerns how the subnet observes useful work and turns it into rewards. The relevant distinctions include whether truth is deterministic, synthetic, private, delayed, or judgment-based; whether validators rerun submissions; whether miners commit before evaluation; whether duplicate contributions share or lose value; whether scores are absolute or rank-based; and whether past performance persists through reputation or incumbent state.

Keeping the two dimensions separate prevents misleading comparisons. ReadyAI, a semantic-extraction subnet, is close to CLAIMS in the transformation from unstructured text to structured records, yet it does not produce a versioned scientific claim–evidence graph. Apex, an objective optimization subnet, is distant in product space, but its hard gates, common execution conditions, and concentrated frontier rewards show what becomes feasible when correctness is inexpensive to recompute. Minos, a genomic benchmarking subnet, is farther from the CLAIMS customer object than ReadyAI but closer in hidden scientific evaluation and productive benchmark design.

Figure 1 places these examples on the two dimensions without collapsing them into one score.

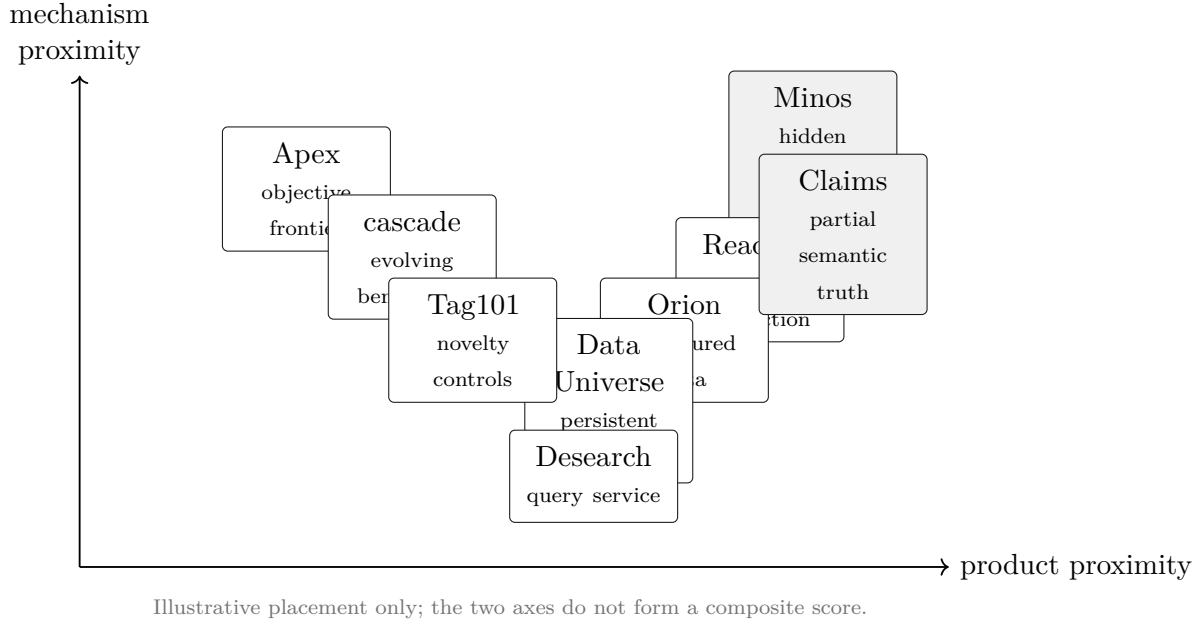


Figure 1. Product proximity and mechanism proximity can diverge. The comparison set includes subnets close to Claims on either dimension rather than only those sharing a category label.

The map is illustrative rather than ordinal. A subnet can be an informative comparator because it is close on either axis, even when the other axis places it far from CLAIMS.

1.2 Questions and scope

With those dimensions in view, the comparison asks four questions.

1. What economic commodity does CLAIMS produce, and which subnets create plausible substitutes or complements?
2. Which truth structures and validation technologies appear across related subnets?
3. Which reward, memory, commitment, and duplicate-control mechanisms recur across the ecosystem?
4. What changes when those components are applied to semantic records for which complete ground truth is expensive or unavailable?

No single ranking can answer all four questions. A mechanism can work well for an objective benchmark and fail in semantic curation. A product can be commercially adjacent while offering little guidance about verification. Each comparison therefore identifies the economic problem that a shared component solves and the condition that limits its transfer.

The paper first fixes the CLAIMS product and mechanism. It then compares scientific, data-construction, retrieval, truth, reward, and validator architectures before placing the closest relatives on a common template and separating operating evidence from token-market evidence.

1.3 Findings at a glance

Five findings organize the comparison. Each identifies a strong relationship and the qualification that keeps the relationship from becoming an equivalence.

Table 1. Five findings that locate Claims in the comparison set.

Finding	Comparative basis	Limiting condition
No featured subnet is an exact twin	No other featured system combines source-grounded scientific records, partial semantic truth, item-level canonicalization, layered verification, and prospective family-level marginal information.	The individual components all have precedents elsewhere.
Minos is the closest composite mechanism relative	Both systems combine scientific work, competition over private methods, hidden evaluation, and an output that can improve later production.	Minos has a more synthetic and scalar truth structure, and its complete production path is not independently verified in the available operating record.
Data subnets supply several neighboring product forms	Persistent inventories, structured training data, semantic tags, and synthetic datasets show distinct ways to create reusable data.	Those products do not automatically provide claim–evidence relations, source fidelity, or reversible canonicalization.
CLAIMS combines familiar mechanisms under harder evaluation conditions	Hidden tests, replay, commitment, concentrated ranks, duplicate penalties, contribution histories, and frontier rewards all appear elsewhere.	Partial semantic verifiability prevents several objective-benchmark rules from transferring unchanged.
Current operation differs materially from prospective assurance	V0 includes structured extraction, Bronze comparison, origin-blind Silver adjudication, multiplicative scoring, and concentrated ranks.	Gold, Anti-Gold, sampling, miner judging, behavioral families, and canonical production ingestion remain prospective.

Taken together, the findings place CLAIMS near data construction in product space and near hidden scientific evaluation in mechanism space. Its distinctive burden is to integrate those elements when the evaluator cannot cheaply observe the complete desired record.

1.4 Operating-status convention

Mechanism design and live operation are separate facts. The comparison therefore distinguishes an active mechanism from a published design, a pilot or testnet, and an unresolved slot. Most

featured subnets are recorded as active. Where active operation is not established, the relevant discussion is confined to the published mechanism and reported data.

Operating status is also separate from earning-miner activity. A mechanism may be identifiable as active even when a dated snapshot shows no miners receiving earnings. That distinction remains in force throughout the paper and matters especially when the next section defines the current CLAIMS mechanism apart from its broader proposed architecture.

The appendices preserve the full supporting record. Appendix A gives the subnet comparison matrix; Appendix B records short comparator profiles; Appendix C separates current operation from prospective directions; Appendix D expands the dated market snapshot; and Appendix E defines the paper-specific terms used throughout.

2 Claims as the Reference Architecture

A comparison of neighboring systems requires a fixed account of what CLAIMS produces and which parts of its design are already operating. The central object is a persistent scientific record rather than a free-form answer. Its substantive content, supporting evidence, provenance, and revision history remain available for later inspection.

2.1 The product: machine-readable scientific evidence

A CLAIMS record represents a well-scoped scientific proposition together with the context needed to understand and audit it. Depending on the paper and extraction schema, a record can include:

- the claim statement, including its modality, direction, scope, and qualifications;
- the paper, section, passage, table, figure, or supplement from which the claim is derived;
- the observation, estimate, experiment, or analysis that bears on the claim;
- populations, interventions, comparators, outcomes, quantities, units, intervals, and uncertainty;
- normalized entities and relations among claims, evidence, methods, and sources;
- extraction, adjudication, and verification provenance;
- the current record state, earlier versions, and unresolved alternatives.

Persistence changes the economic object. A search response serves one request, whereas a durable claim–evidence record can support later queries, analyses, training tasks, and corrections. Its value therefore depends on more than initial accuracy. Stable identity, continued source access, visible lineage, and reversible revision all become part of the product.

2.2 Paper content, accepted record, and scientific truth

Scientific curation must keep three objects separate. For paper p , let X_p denote the content that the extraction schema asks the system to recover. Let F_p denote the record that the system currently accepts, and let W_p denote the underlying truth of the scientific proposition. In general,

$$\boxed{X_p \neq F_p \neq W_p.}$$

The first objective is to make F_p a faithful structured representation of X_p . A record can accurately report a study whose real-world conclusion is later overturned. Conversely, later evidence can

overturn a scientific claim without showing that the original extraction was wrong. This boundary separates CLAIMS from forecasting systems settled by later outcomes and from benchmark systems whose answer key directly represents the target truth.

2.3 Four information layers

The extraction object combines four layers, and each layer places a different burden on the evaluator.

1. **Coverage and source structure.** Sections, spans, tables, figures, identifiers, and document anchors.
2. **Grounded facts.** Values, units, sample sizes, intervals, interventions, comparators, and outcomes that can often be checked directly against the source.
3. **Semantic judgments.** Claim identity, relation type, direction, modality, causal status, evidence role, and equivalence.
4. **Open or higher-order judgments.** Assessments for which the source does not supply a complete immediate answer key.

Inexpensive deterministic checks are strongest in the first two layers. The third relies more heavily on semantic adjudication. The fourth may require trusted review, later evidence, carefully structured consensus, or an explicit unresolved state. One aggregate accuracy score can conceal a failure in one layer behind strength in another, so the architecture keeps these burdens visible.

2.4 Bronze, Silver, Gold, and Anti-Gold

The four tiers separate economical production, stronger machine adjudication, trusted real cases, and controlled synthetic errors.

Bronze is the economical reference extraction. It is broad enough to provide a useful starting point and inexpensive enough to run at scale. Economically, Bronze prevents repeated payment for work below a known baseline, reduces the cost of comparing heterogeneous methods, and directs miners toward residual errors and omissions.

Silver is the stronger machine-curated record produced through comparison, evidence-aware adjudication, and canonicalization. Silver can use more capable models, richer context, repeated passes, and targeted verification that would be too expensive to apply everywhere. It preserves valid Bronze content, accepts verified miner improvements, merges equivalent representations, and retains ambiguity when the evidence does not support one answer.

Gold is authoritative or otherwise trusted information from real or known-answer cases. It can resolve high-value disputes, evaluate Bronze and Silver, calibrate judges, gate releases, and reveal errors outside the current challenge bank. Gold remains sparse because authoritative review is costly.

Anti-Gold consists of hidden known-answer tasks created through controlled transformations. The system can delete a qualifier, perturb a number, swap a comparator, change a source link, or reverse a claim–evidence relation. Because the planted transformation is known, the platform can score whether a producer detects it. Anti-Gold can estimate reliability and discipline effort without requiring a complete expert extraction for every test.

Synthetic verification nevertheless has a boundary. An error class is *represented* only when the system can generate a valid challenge, conceal it, score the response, and transport performance back to ordinary production. Errors outside that set remain *latent*. Better Anti-Gold can replace routine Gold on represented classes. It can also make exploratory Gold more valuable because discovering a latent class creates a new domain for future synthetic supervision. Expertise consequently moves toward the boundary of what the system does not yet know how to test.

Figure 2 shows how the current production path and the prospective assurance layers fit together.

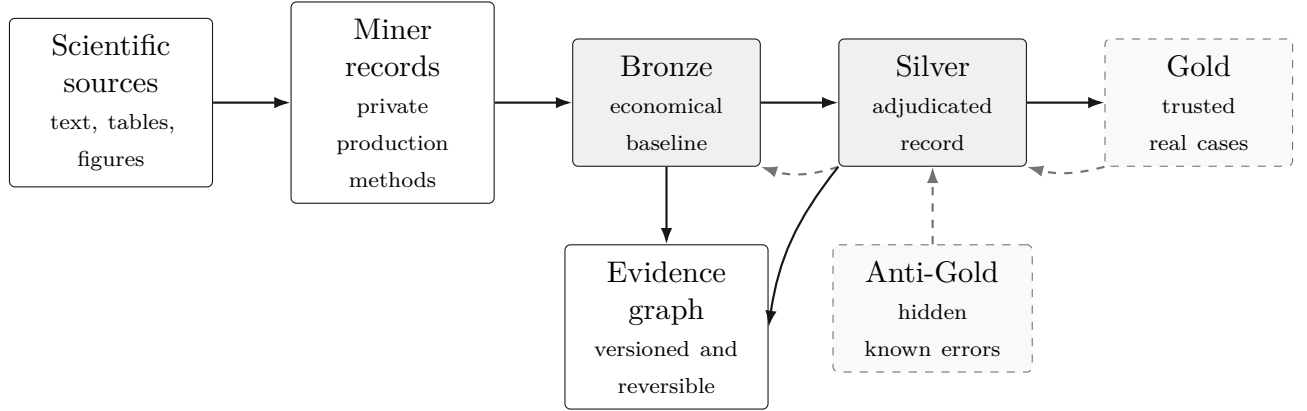


Figure 2. The Claims architecture separates current production from prospective assurance. Solid boxes show current v0 objects; dashed boxes show proposed components. Review moves records toward stronger verification, while accepted corrections improve lower-cost production.

The architecture moves records toward stronger verification while allowing accepted corrections to improve lower-cost production. The dashed components add assurance only if they supply information that the current path does not already contain.

2.5 Canonicalization and reversible lineage

Literal text is not the economic unit. Two miners can express the same scientific proposition in different words, one claim can refine another without contradicting it, and a paper can contain either two distinct claims that a shallow system merges or one claim that another system splits. Canonicalization maps equivalent expressions to shared items while preserving material differences in polarity, modality, numerical value, population, comparator, and evidence.

Stable item identity serves both data quality and incentives. It lets the system attribute accepted contributions, share credit for genuine corroboration, avoid repeated payment for superficial paraphrases, and calculate which information would disappear if a strategy family were removed. When the evidence remains ambiguous, the record preserves that ambiguity instead of forcing a false single answer.

Record status and participant score also remain separate. A strong miner does not automatically certify every submission, and a low-scoring miner can still supply one item that later receives independent support. The database therefore stores evidence at the item level, while the reward mechanism estimates the usefulness and reliability of participant behavior.

2.6 Current v0 mechanism

The active mechanism implements the production and adjudication core, but it does not yet implement the full assurance architecture.

Current mechanism. V0 is an active calibration mechanism. It uses structured scientific extraction, deterministic structural and grounding checks, a versioned Bronze reference, origin-blind evidence-aware adjudication, Silver canonicalization, multiplicative paper scoring, arithmetic batch aggregation, and a concentrated top-five payout. It does not contain an incentive-bearing Gold tier, an Anti-Gold challenge lane, a post-submission sampling stage, a miner-judge role, behavioral-family scoring, a reputation transformation, or automatic canonical production ingestion.

Table 2 records the operating controls that shape the current comparison.

Table 2. Current v0 operating controls and their comparative significance.

Control	Current value	Comparative significance
Papers per batch	40–50	Broad multi-paper production rather than one isolated benchmark item
Target miners	Approximately 10	Selective competition within a larger registered population
Miner response window	30 minutes	A hard constraint on model, parsing, repair, and upload choices
Round cadence	6 hours	Four rounds per day, with validation and recovery between deadlines
Paper score	coverage $\times$ diagnostic quality $\times$ adjudication quality	Complementary dimensions; weakness in one can materially reduce the total
Batch score	Arithmetic mean over scoring-eligible papers	No v0 quality gate, minimum statistic, or reputation average
Payout	70/16/8/4/2 percent for ranks 1–5	Strongly concentrated relative-rank competition
Substantive validator	One owner-operated validator	Current evaluation is not substantively decentralized
Canonical ingestion	Off	V0 outputs remain calibration artifacts rather than automatic production truth

For each paper, the validator hydrates a signed artifact, runs structural and grounding checks, creates or retrieves Bronze, and projects claims into comparison candidates. Source overlap and semantic retrieval open bounded candidate pairs. Anonymous candidates then receive evidence-aware adjudication, and accepted candidates are canonicalized into Silver units.

The paper score has three multiplicative components. Coverage measures how much of the final scored record the miner supplied. Diagnostic quality measures structural, grounding, and rigor findings. Adjudication quality penalizes invalid additions. The validator averages paper scores across the scoring-eligible batch and then maps relative rank into the fixed payout schedule.

That final transformation turns a continuous score into a discontinuous economic contest. A small improvement near a paid-rank boundary can have large reward value, while the same improvement below fifth place may have no immediate payout effect. This feature becomes important in comparisons with winner-take-all benchmarks, proportional contribution ledgers, and long-memory reputation systems.

2.7 Prospective v1 extensions

The proposed v1 architecture changes assurance and participant roles rather than merely increasing throughput.

Prospective extension. V1 adds trusted Gold, hidden Anti-Gold, post-commit sampling, separate extractor and judge tasks, commit–reveal panels, reliability weighting, behavioral-family treatment, marginal contribution, verification-tiered canonical ingestion, frozen reference releases, independent replay, and additional substantive validators. Each component becomes incentive-bearing only after its measurement and failure conditions are defined.

These additions introduce exogenous calibration, test evaluator reliability, make identity treatment less dependent on UIDs, and turn accepted content into a versioned production asset. They also create new failure modes. Synthetic tests can become recognizable; family clustering can merge independent methods or split one copied method; sampled evaluation can increase score variance; and multiple validators can remain correlated if they rely on the same model family or private service.

The distinction between v0 and v1 is therefore substantive. The current mechanism supplies the reference point for live comparisons, while prospective components are evaluated only as proposed extensions whose value depends on the additional information they can produce.

3 Scientific and Technical Comparators

Scientific subnets do not produce one common commodity. Some solve objectively scored computational problems, some optimize engineering artifacts, some propose molecules or biological designs, some infer hidden states from scientific measurements, and some generate synthetic data. CLAIMS instead represents the content and evidence of existing research. That position in the scientific process changes both the customer object and the evaluator’s access to truth.

3.1 Where the products enter the scientific process

A simplified production chain helps separate neighboring outputs:

sources and observations → structured evidence → analysis and synthesis
→ candidate generation → experiments and outcomes.

CLAIMS operates mainly at the structured-evidence stage. Minos infers biological variants from sequencing reads. NOVA proposes candidate molecules and nanobodies. NIOME generates synthetic biological datasets. Apex, CliqueAI, and ChipForge solve formal or engineering problems under common benchmarks. These systems can serve the same scientific organizations without producing interchangeable outputs.

Figure 3 locates the products along that chain.

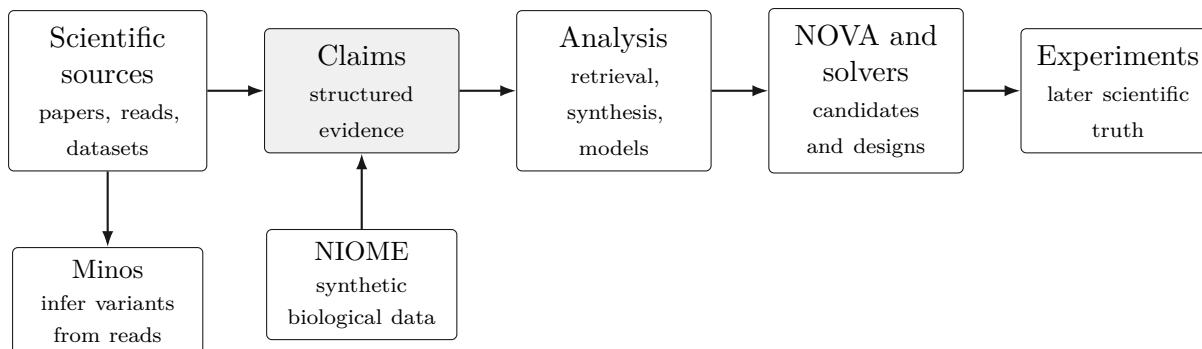


Figure 3. Claims occupies the evidence-representation stage of a broader scientific production chain. Neighboring scientific subnets can therefore complement Claims even when their validation mechanisms look similar.

The placement explains why product complementarity can coexist with mechanism similarity. A downstream scientific-design contest can use evidence curated upstream even when both systems rely on hidden evaluation.

3.2 Minos: the closest composite comparison

Among the scientific comparators, Minos comes closest to the combined CLAIMS mechanism because both systems standardize a scientific task, let miners keep production methods private, and evaluate the resulting work against information withheld by validators.

Operating-status note. Minos is treated as a published current design with reported earning-miner activity and market data. The complete production task-to-weight path, current validator code, exact metric blend, and allocation schedule are not independently established in the available operating record. The discussion therefore concerns its published mechanism and reported data rather than a fully reproduced live implementation.

Minos organizes genomic variant calling as a hidden-test performance contest. Miners tune or develop pipelines that infer single-nucleotide variants and short insertions or deletions from noisy sequencing reads. Validators rerun those configurations on synthetic sequencing cases whose inserted variants remain hidden. False positives and missed variants lower performance, while invalid or non-executable submissions fail before comparison.

The shared architecture is substantial. Both systems use a concrete scientific task rather than generic inference. Both standardize the task and output while leaving methods private. Both benefit from blind evaluation and place substantive work inside the validator instead of assuming that base-layer consensus establishes correctness. Both can also leave behind reusable scientific data and difficult examples rather than only a leaderboard.

The answer key separates them. A synthetic genome can contain a private set of inserted mutations against which every call is scored. Scientific extraction rarely permits an equally complete construction. The evaluator can inject a wrong number or relation, but ordinary records also fail through omissions, over-merging, lost qualifications, unsupported causal language, incorrect

evidence roles, and ambiguity. Some of those failures can be synthesized; others remain latent until verification of real cases discovers them.

A reliable scalar benchmark can support aggressive rank concentration. A partly verifiable semantic task instead requires several score dimensions, selective adjudication, abstention, explicit unresolved states, and independent tests of the evaluator. Minos demonstrates the value of hidden scientific truth and private-method competition. CLAIMS adapts that architecture to a domain in which hidden truth is useful but incomplete.

3.3 NOVA: scientific generation rather than evidence curation

NOVA clarifies the difference between producing a scientific design and representing published evidence. It is recorded as active and runs target-specific tournaments for molecules and nanobodies. Miners commit encrypted candidates before reveal. Validators reject invalid or unrevealed designs and rank the remaining candidates with pinned models of structure, binding, novelty, and diversity. Commitment prevents adaptation after competitors or evaluation information become visible, while duplicate controls reduce the return to near-identical candidates.

NOVA and CLAIMS may serve overlapping life-science users, but they create value at different stages. NOVA proposes prospective designs. CLAIMS represents the published evidence that can inform target selection, compare prior findings, and help interpret later experiments. NOVA’s operative truth is predicted biological performance under maintained models rather than eventual experimental success. The first truth target for CLAIMS is fidelity to the source rather than the real-world validity of the paper’s conclusion.

The mechanism comparison still transfers one useful primitive. NOVA uses commit–reveal to protect a design contest; the prospective CLAIMS judge lane uses commitment to prevent copying or adaptation during evaluation. Both bind an action before later information arrives, even though the hidden action they deter differs.

3.4 NIOME: synthetic scientific data

NIOME is recorded as active and rewards biologically credible synthetic CRISPR datasets. Validation moves through syntax, statistical plausibility, validator-only simulation, internal consistency, and comparison with held-out real distributions. A dataset can satisfy the formal schema and still fail because its biological behavior is implausible.

This mechanism separates dataset realism from document fidelity. NIOME asks whether generated records behave like plausible biology. CLAIMS asks whether records faithfully represent identifiable scientific sources and whether their relations survive adjudication. Both use staged validation and private evaluator components. NIOME’s main truth boundary lies in biological models and held-out distributions; the CLAIMS boundary lies in the paper corpus, source parsing, Bronze, semantic judges, Gold, and canonicalization.

NIOME also shows why raw volume has little value without validation. The same principle applies to CLAIMS: generating many plausible statements is not useful when they omit required content, lack source support, or add invalid claims.

3.5 Objective scientific and engineering tournaments

Apex, CliqueAI, and ChipForge, all recorded as active, occupy the opposite end of the truth spectrum. Their outputs can often be checked against a complete formal or engineering test.

Apex runs owner-defined optimization contests. Miners submit executable solutions, validators run them under common conditions, and the best valid result holds the allocation until displaced. CliqueAI supplies graph instances and verifies whether a proposed clique is valid and maximal; eligible solutions can then be compared by size, difficulty, and diversity. ChipForge tests hardware designs against concealed functional cases and, where configured, engineering objectives such as area, timing, or power.

Because invalid work is inexpensive to reject, the evaluator can often reproduce the object that determines reward. Winner-take-all or frontier-ratchet mechanisms are then credible: an overtaking result is observable, and a copied solution can be discounted without an open-ended semantic judge.

CLAIMS cannot import that structure unchanged. A scientific record can be syntactically valid and semantically wrong. Two different records can both be defensible, and the best final record for one paper can combine contributions from several miners. One complete submission need not dominate every item. CLAIMS therefore constructs Silver item by item even though v0 still concentrates payout at the batch-rank level.

Table 3 summarizes the product, truth source, and mechanism distinction for the scientific comparison set.

Table 3. Scientific and technical comparators.

Subnet	Primary object	Main truth source	Mechanism backbone	Decisive difference from CLAIMS
CLAIMS	Source-grounded claim–evidence records	Direct source checks and semantic adjudication; prospective Gold and Anti-Gold	Verified contribution plus rank tournament	Reference case
Minos	Variant-calling pipelines and synthetic genome outputs	Hidden synthetic mutations and validator replay	Benchmark tournament	More complete synthetic answer key; implementation path only partly verified
NOVA	Molecules and nanobodies	Pinned predictive models and validity gates	Benchmark tournament	Scientific design rather than evidence representation
NIOME	Synthetic CRISPR datasets	Staged plausibility, simulation, and private distributions	Verified contribution ledger	Dataset realism rather than source fidelity
Apex	Executable optimization solutions	Common objective benchmark	Winner-take-all frontier	Inexpensive scalar verification
CliqueAI	Maximal-clique solutions	Formal graph validity and comparative solution quality	Benchmark tournament with history	Objective validity and long-memory reputation
ChipForge	Hardware designs	Hidden functional tests and engineering tools	Frontier ratchet	Objective qualification and incumbent displacement

Across these systems, scientific subject matter alone says little about the right incentive design. The decisive variable is how fully the validator can observe the desired output, a distinction that becomes even sharper among data-construction systems.

4 Data Construction and Semantic Extraction

The label *data subnet* hides several distinct products. Collection, inventory, tagging, training-data construction, semantic extraction, and canonical knowledge representation each transform source material to a different degree. The closer a product moves toward reusable scientific meaning, the more its evaluator must judge relations rather than only availability or format.

4.1 A hierarchy of data products

Figure 4 organizes the comparison by the transformation applied between source and output.

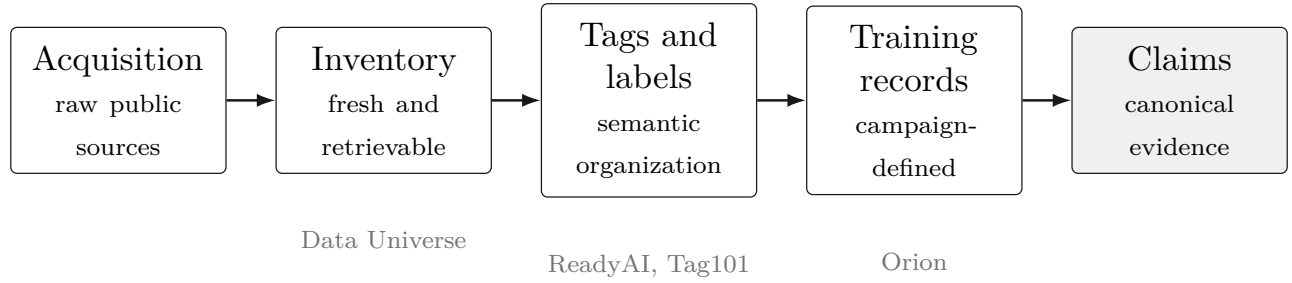


Figure 4. Data products differ by how much structure they add to source material. Claims sits downstream of acquisition and labeling because it adds scientific relations, canonical item identity, and reversible lineage.

The ladder is not a ranking of economic value. Fresh raw data can be more valuable than heavily transformed records for some uses. Each stage instead solves a different production problem and therefore requires a different evaluator. CLAIMS occupies the final stage shown because it adds claim identity, evidence relations, canonicalization, and reversible lineage.

4.2 Data Universe: persistent source inventories

Data Universe is recorded as active with broad earning-miner activity. It rewards miners for maintaining useful, verifiable inventories of public internet data. Value depends on more than volume: freshness, demand, duplication, and credibility inferred from spot checks all matter. Validators revisit original sources and can reduce the value of a miner’s broader inventory after failed verification.

Persistence creates the strongest connection to CLAIMS. Both systems build an asset that remains useful after the scored task, require continuing access to sources, control duplicates, and direct miners toward demanded content rather than undifferentiated volume.

Semantic depth separates them. Data Universe mainly evaluates whether a record was collected, remains available, is fresh, is relatively scarce, and matches current preferences. CLAIMS must also determine whether a paper’s propositions, evidence, quantities, qualifications, and relations have been represented correctly. A failed spot check can update inventory credibility, but miner credibility alone cannot establish that two scientific sentences are equivalent or that a table result supports a specific causal claim.

4.3 Orion: structured training-data production

Orion is recorded as active, although the dated snapshot classifies it as pre-market with no earning miners. It operates a campaign-based market for structured training data. The operator issues signed assignments that specify the requested data, schema, storage path, and evaluation samples. Invalid, stale, duplicate, unavailable, or unplayable work receives no value. Accepted contribution depends jointly on verified quality, useful output volume, and continued availability.

The workflow is one of the closest production analogs to CLAIMS. Both systems assign bounded source material, prescribe a structured artifact, use signed outputs, apply hard validity gates, inspect hidden samples, and retain information needed for audit. Both reward usable production rather than nominal compute.

The final commodity differs. Orion produces campaign-defined training material. CLAIMS is intended to produce a durable scientific evidence graph whose items retain identity beyond the batch that created them. It must reconcile competing representations, preserve evidence relations, and allow later corrections to supersede a record without erasing earlier states. Canonicalization and lineage are therefore product features rather than optional audit metadata.

4.4 ReadyAI: semantic extraction under reference evaluation

ReadyAI is recorded as active with broad earning-miner activity. It converts text into tags, annotations, and other semantic structure. Validators compare miner outputs with reference records through semantic similarity and judged coverage or correctness. The mechanism favors several distinct relevant labels and limits the value of one strong match surrounded by weak or repetitive tags.

ReadyAI is close to CLAIMS in both product and evaluation. Each turns unstructured text into structured semantic objects and depends on coverage, correctness, uniqueness, and an external evaluation stack. Each also faces the risk that the reference record and judge become the effective truth source.

CLAIMS adds an interdependent relational object. A claim is more than a tag: its meaning depends on qualifications, evidence, source locations, populations, comparators, quantities, and relations to other claims. A miner can identify the right topic while attaching the wrong evidence or overstating the paper’s modality. The evaluator therefore needs source-grounded candidate packets and evidence-aware adjudication, not only semantic similarity between labels.

4.5 Tag101: novelty and repeated semantic output

Tag101 is recorded as active with broad earning-miner activity. It rewards timely topic tags attached to eligible social posts. Value depends on fit with an active cluster, the value of the underlying post, and semantic distance from the miner’s own recent tags. Averaging across tags and penalizing repetition reduce the return to raw volume and reused tag sets.

The comparison isolates redundancy. CLAIMS must also distinguish new information from superficial variation. Two miners can independently corroborate one claim, one miner can submit several paraphrases, and one strategy can appear across several identities. Tag101 addresses repetition through semantic history at the tag level. The prospective CLAIMS design extends the same economic logic to canonical scientific items and behavioral strategy families.

That logic cannot transfer mechanically. Repeated scientific agreement can be valuable when it comes from independent methods. A duplicate penalty that ignores independence can suppress legitimate corroboration. CLAIMS therefore needs bounded corroboration value and a separate marginal-contribution calculation rather than treating all similarity as spam.

4.6 cascade: an evolving data benchmark

cascade is recorded as active. It evaluates deterministic synthetic-data generators by training a common forecasting model under common budgets and testing whether the resulting model improves on private real-world series. Inexpensive heats screen candidates before a paired bootstrap comparison determines whether a challenger has credibly displaced the current king. Immutable generator artifacts, future-block seeds, and public receipts support replay.

Both cascade and CLAIMS contain a compounding loop. A strong cascade generator can improve downstream training data. A verified CLAIMS correction can improve Silver and, through Silver, Bronze. In each system, successful output changes the baseline that future competitors face. Release control and holdouts therefore carry substantive weight rather than serving as administrative details.

The evaluation object again differs. cascade can fix a downstream model, budget, and private test series, then compare generator performance statistically. Reducing CLAIMS data quality to one downstream model would narrow the product to that model’s current blind spots. CLAIMS instead requires several quality dimensions, held-out Gold, random re-audit, and explicit protection against a generator, producer, and adjudicator sharing one semantic failure.

Table 4 places these neighboring products on a common set of dimensions.

Table 4. Data-construction and semantic-extraction comparators.

Subnet	Data product	Validation emphasis	Persistence	Extension required for CLAIMS
Data Universe	Fresh public-web inventory	Authenticity, freshness, scarcity, duplication, and credibility	Persistent inventory	Semantic claim identity and evidence relations
Orion	Campaign-defined training records	Signed assignments, hidden replay, quality, volume, and availability	Persistent campaign output	A canonical scientific graph across batches
ReadyAI	Tags and semantic annotations	Reference similarity and judged coverage and correctness	Recurring semantic records	Source-grounded relational records
Tag101	Timely topic tags	Cluster fit, post value, novelty, and repetition history	Time-windowed contribution history	Independent corroboration and item-level marginal information
cascade	Synthetic forecasting data	Downstream model improvement on private series	Champion generator and evolving baseline	Several evidence dimensions and external Gold
CLAIMS	Claim–evidence records	Structure, grounding, semantic adjudication, and accepted improvement	Versioned evidence graph	Reference case

The data comparison places CLAIMS downstream of acquisition, inventory, and labeling without implying that downstream products always dominate upstream ones. The next comparison asks when a user may bypass preconstructed records altogether and rely on retrieval at the moment of need.

5 Search and Retrieval: Substitute or Complement?

At the user interface, search and retrieval are the closest visible alternatives to a preconstructed evidence graph. A user can ask a retrieval system to find papers and synthesize an answer without first paying to curate a standardized record. The relationship changes when the same sources are used repeatedly, when findings must be normalized across papers, and when provenance and correction histories matter.

5.1 Desearch: a query-time service

Desearch is recorded as active with broad earning-miner activity. It rewards useful search and research responses based on retrieved content, summaries, social-media results, corroboration, validity, and response performance. Moving miner scores turn continuing query traffic into a service competition.

Desearch and CLAIMS can both deliver grounded scientific information, but they produce it at different times. Desearch retrieves and synthesizes when a request arrives. CLAIMS curates in advance and stores standardized evidence objects that later queries can reuse. Query-time production economizes on upfront curation when demand is sparse or idiosyncratic. A persistent asset economizes on repeated reconstruction when the same papers and claims support many applications.

5.2 rec4ll: announced retrieval and persistent memory

rec4ll matters to the comparison at the level of product architecture, although its current mechanism cannot support a parallel reward analysis.

Operating-status note. The operating state of rec4ll is unresolved. Its current benchmark, truth source, eligibility rules, aggregation, routing incentives, and score-to-weight path are not established. The comparison refers only to its announced retrieval and persistent-memory architecture and to the dated market data associated with the slot.

The announced design combines embedding, vector search, retrieval, and language-model capacity. That architecture is relevant because CLAIMS records can become retrieval units, while retrieval systems can provide the discovery interface around the evidence graph. The unresolved mechanism limits the comparison to this product-level relationship.

5.3 Conditions for substitution

Query-time retrieval can substitute for CLAIMS when a user needs a one-off answer, only a few papers matter, cross-study normalization is unnecessary, and a persistent audit or correction history has little value. Under those conditions, curating a large evidence graph before demand appears may impose costs that later use does not recover.

5.4 Conditions for complementarity

The systems become complements when structured records reduce the cost or raise the quality of later search. CLAIMS can supply normalized entities, claim-level retrieval units, evidence links,

source spans, qualifications, and reusable provenance. Search can discover new sources, route users to relevant records, and fill gaps outside the curated corpus.

Table 5 makes the production-time tradeoff explicit.

Table 5. Query-time retrieval compared with persistent evidence infrastructure.

Dimension	Query-time retrieval	CLAIMS evidence infrastructure
Unit produced	Ranked sources and a synthesized answer	Canonical claim–evidence item
Production time	When a request arrives	Before and during repeated use
Persistence	Often specific to one request	Versioned and reusable
Grounding	Sources attached to a response	Item-level source spans and evidence relations
Correction	A new answer on a later request	Explicit supersession and lineage
Economy	Payment for serving and current retrieval	Payment for curation, assurance, and asset maintenance
Strongest use	Sparse, changing, or idiosyncratic questions	Repeated cross-paper comparison and machine reuse

Search can therefore become a distribution layer for a persistent evidence asset, while CLAIMS can become a structured substrate for better retrieval. Whether the relationship is substitution or complementarity depends on the pattern of use rather than on the interface alone.

6 Truth Structures and Verification

The decisive mechanism difference across subnets is usually the evaluator’s access to truth, not the model a miner uses. The same reward rule can work well when correctness is inexpensive to recompute and fail when the desired output is only partly observable. CLAIMS belongs to the latter case, so its comparisons must begin with the source and cost of an answer key.

6.1 A taxonomy of truth sources

Five truth structures recur across the comparison set.

Deterministically verifiable.

Correctness or feasibility can be recomputed cheaply. Formal graph validity, code execution, and functional hardware tests are examples.

Synthetic ground truth.

The evaluator constructs hidden tasks with known answers, as with inserted genomic mutations or matched generated examples.

Partial verification.

Some properties are objective, while others require semantic judgment, selective trusted truth, or an explicit unresolved state. CLAIMS belongs mainly to this category.

Delayed outcomes.

Truth becomes observable only after markets, games, deployments, or later events resolve.

Open judgment.

No timely complete answer key exists, so consensus, peer prediction, external review, or cautious abstention may be required.

Figure 5 arranges these structures by the cost of obtaining a complete answer key.

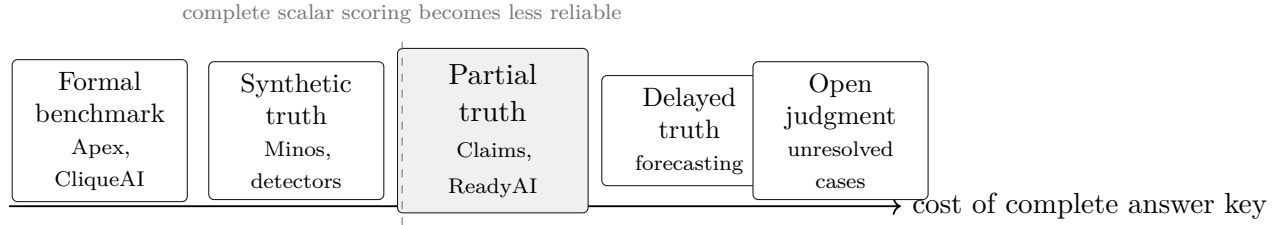


Figure 5. The cost and completeness of an answer key constrain mechanism design. Claims combines inexpensive direct checks with costlier semantic judgment and trusted verification.

As the answer key becomes more expensive or incomplete, a single scalar score becomes less reliable. The mechanism must then combine several evidence sources and expose unresolved cases rather than hiding them inside an average.

6.2 Hidden tests and known-answer tasks

Hidden tests reduce the return to effort aimed only at recognizable evaluation cases. They can prevent miners from reserving careful work for calibration items, permit repeated capability measurement, and refresh evaluation when public benchmarks become saturated.

The implementations target different forms of adaptation. Minos hides synthetic genomes. BitMind refreshes media data and uses private holdouts. ByteLeap commits a larger result set before validators sample proofs. NOVA commits candidates before reveal and evaluation. In each case, the participant acts before learning which item, sample, competitor, or evaluation draw will determine reward.

Informational timing is the common mechanism, but the deviation being controlled differs: benchmark overfitting, copied voting, selective computation, or retrospective candidate adaptation. Hidden evaluation therefore transfers only after the relevant hidden action has been identified.

6.3 Synthetic truth and synthetic falsity

Some tasks permit a complete synthetic answer key. A validator that inserts known variants into a synthetic genome knows which calls should be present. A detector benchmark can pair human and generated media with known class labels. Other tasks make it cheaper to generate something known to be false than to construct the unique correct full answer.

CLAIMS often has this asymmetry. A validator can create a record that is wrong with high confidence by changing a source span, number, unit, comparator, polarity, qualification, or relation. The validator need not know the uniquely correct full record to know that the planted transformation is invalid.

That asymmetry makes Anti-Gold scalable, but it also limits what Anti-Gold establishes. Detecting a wrong number does not necessarily reveal the right number, and rejecting an invalid relation does not identify the best complete semantic representation. Synthetic falsity is therefore a verification instrument rather than a substitute for all truth production.

6.4 The four Anti-Gold conditions

Anti-Gold has value only when four conditions hold.

Validity.

The planted transformation must actually create a known error. Binary fields, synonyms, equivalent units, and already-wrong source records can make careless mutations correct or semantically neutral.

Transport.

Performance on synthetic errors must predict or induce performance on organic errors. Easy artificial mistakes can overstate ordinary extraction quality.

Concealment.

Miners must not reliably identify the challenge as synthetic. Recognizable templates reward challenge detection rather than routine care.

Scoreability.

The output contract must make blanket rejection, indiscriminate editing, and other trivial strategies unprofitable.

Coverage adds a fifth requirement. Repetition within a fixed mutation basis improves measurement only for represented error directions; it cannot reveal a latent class outside that basis. Genuine review remains distinct because it can discover a failure that the system did not yet know to test.

6.5 Routine and exploratory Gold

The distinction between represented and latent classes changes how better synthetic supervision affects trusted review. On represented classes, Anti-Gold and routine Gold can serve the same monitoring objective. Better Anti-Gold therefore reduces the need to repeat expert checks of familiar failures.

At the frontier, real verification plays a different role. It can reveal a missing error class and convert that discovery into future synthetic capacity. As a result, better Anti-Gold raises the continuation value of exploratory Gold even while replacing routine Gold inside the installed challenge bank.

The composition of Gold shifts accordingly. Routine review moves away from familiar errors, while exploratory review becomes more valuable at a responsive frontier. Total Gold can rise or fall depending on whether the leverage from discovering new classes outweighs substitution across represented classes. Once exploration reaches a limit, the frontier response stops and the remaining routine substitution reduces total Gold.

This distinction is especially relevant to CLAIMS because model progress can make existing mutations inexpensive while new semantic failures continue to appear. A mature challenge bank does not remove the need for random review of apparently undisputed content, disagreement cells outside represented classes, or external audits of shared blind spots.

6.6 Commitment and post-commit sampling

Commitment mechanisms bind work before later information arrives, but they solve several different timing problems.

- **Commit–reveal** binds a candidate or vote before later information is revealed.
- **Immutable artifacts** prevent code, data, or generators from being modified after selection.
- **Hidden holdouts** prevent direct optimization against the scored cases.
- **Post-commit sampling** requires complete work before the validator chooses what to inspect.

ByteLeap is recorded as active and provides a direct analog for post-commit sampling. Idle workers commit a larger proof set before validators inspect random samples. The prospective CLAIMS design applies the same timing to papers or items: miners process the assigned workload, commit before selection, and face a positive probability that any item will be audited. Current v0 does not use this mechanism.

Commitment lowers the return to selective effort only when the committed universe is complete, the sample is representative, and the work can be replayed. Those conditions are part of the mechanism rather than implementation details.

6.7 Delayed truth

Hah is recorded as active, although the dated snapshot classifies it as pre-market with no earning miners. Its forecasting mechanism evaluates probabilities after sports outcomes and closing lines become available. Performance combines calibration, market edge, timeliness, uniqueness, and marginal contribution. Delayed outcomes make it possible to distinguish skill from a plausible narrative.

Scientific replication can supply a related long-run signal, but it is sparse, delayed, selected, and dependent on the field. Replication can help calibrate claims about scientific credibility without serving as the sole near-term test of whether a miner faithfully extracted a paper. CLAIMS therefore keeps source fidelity separate from later scientific settlement.

6.8 Verification and truth production

Expert work has two distinct uses. *Verification* estimates whether a producer performs the desired activity. *Truth production* establishes or corrects the canonical content of a real case. Synthetic tasks can substitute strongly for verification when validity, transport, concealment, scoreability, and coverage hold. They generally cannot eliminate truth production.

This separation prevents an overstatement of automation. A system can use many inexpensive known-negative tasks to estimate miner reliability while reserving high-quality real-case review for disputes, new error classes, release gates, and authoritative corrections.

Table 6 summarizes how the principal truth architectures trade scale against completeness.

Table 6. Truth and verification architectures.

Architecture	Representative systems	Main strength	Main limitation
Deterministic benchmark	Apex, CliqueAI, ChipForge	Inexpensive validity checks and reproducible comparison	Applies only to tasks reducible to the benchmark
Synthetic answer key	Minos, BitMind classifiers	Fresh hidden labels and scalable scoring	The generator and holdout define the tested world
Known-negative Anti-Gold	Prospective CLAIMS challenges	Inexpensive tests without a complete positive record	Does not itself produce the correct live record
Partial semantic verification	CLAIMS, ReadyAI, Tag101	Combines direct checks with rich semantic output	Judge calibration and common-mode error
Post-commit sampling	ByteLeap; prospective CLAIMS sampling	Disciplines selective production at lower audit cost	Requires representative samples and replay
Delayed outcome	Hah and related forecasts	Later events eventually settle performance	Outcomes are sparse, delayed, and selected

The comparison rules out a universal verification formula. As answer keys become less complete, assurance must come from several independent signals, and the reward system must preserve the difference between measuring a producer and correcting the record.

7 Competition, Rewards, and Independent Information

A subnet can reward benchmark rank, accepted contribution, delivered service, persistent incumbency, or skill revealed by later outcomes. CLAIMS combines two of these ideas: it constructs a verified contribution object and then applies a strongly concentrated rank schedule. That hybrid creates strong incentives near the frontier, but it can leave useful information below a paid rank with little immediate reward.

7.1 Five mechanism families

The comparison set contains five recurring reward backbones.

Benchmark tournament.

Miners compete on a common task and score. Apex, CliqueAI, Minos, NOVA, and BitMind use variants of this structure.

Frontier ratchet.

A current incumbent or crown persists until a credible challenger displaces it. Cacheon, cascade, and ChipForge illustrate this design.

Verified contribution ledger.

Accepted work accumulates value after eligibility and verification. Data Universe, Orion, ReadyAI, Gittensor, Tag101, and CLAIMS contain this logic.

Metered service.

Rewards follow delivered search, compute, storage, or customer service. Desearch and ByteLeap contain service components.

Delayed-outcome skill market.

Predictions or policies are evaluated after outcomes resolve. Hah illustrates this structure.

No backbone dominates across all settings. The appropriate choice depends on whether contributions are additive, whether one best artifact can serve every user, whether performance should persist, and whether the evaluator can measure quality reliably. CLAIMS uses a hybrid because its accepted record can combine miners' contributions even though its current payout rewards relative batch rank.

7.2 Absolute quality and relative rank

The current CLAIMS score is constructed in two stages. First, coverage, diagnostic quality, and adjudication quality determine an absolute paper score. Paper scores are then averaged within the batch, and relative rank maps that batch score into the top-five payout. The first stage measures performance under the evaluator; the second creates economic scarcity.

Figure 6 shows where the discontinuity enters.

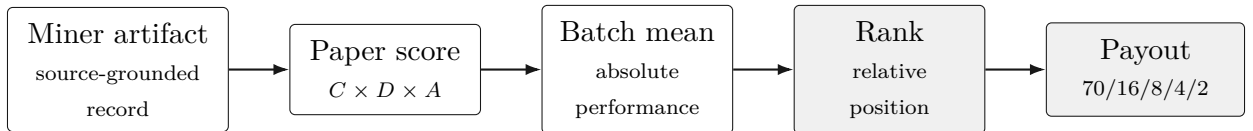


Figure 6. Claims measures quality before concentrating rewards. Continuous paper and batch scores become discontinuous payouts after miners are ranked.

Apex takes the discontinuity farther: the best valid result holds the entire allocation. That rule is attractive when one executable solution can serve the whole task and the benchmark is reliable. The final CLAIMS record can instead combine several miners' accepted items, so winner-take-all treatment at the submission level would discard useful complementarity. V0 retains concentrated batch rewards while constructing Silver item by item.

7.3 Frontier ratchets and incumbent memory

Cacheon is recorded as active and rewards reproducible improvements to validator-selected compute targets. An accepted improvement creates a live claim that decays and is retired when another contribution establishes a new crown. cascade uses a statistically gated king-challenger comparison for synthetic-data generators. ChipForge can keep an active hardware-design winner until a challenger displaces it.

Frontier ratchets conserve attention because the system need not rerank every historical contribution from scratch. They also create path dependence. An incumbent can retain value even when several

alternatives are close, and the replacement margin determines how the design balances stability against contestability.

The main CLAIMS competition currently resets with each batch rather than carrying a record-level incumbent reward. Bronze nevertheless acts as an incumbent information baseline: miners gain by identifying valid content beyond it. A future release can use frontier logic for reference updates while keeping separate batch rewards for current production.

7.4 Contribution ledgers and persistent reputation

Data Universe updates credibility through repeated spot checks. Gittensor maintains a rolling history of accepted open-source work. Orion accumulates accepted production and availability observations. CliqueAI carries long-memory performance across graph tasks. In each case, history reduces noise and makes sustained reliability economically relevant.

CLAIMS v0 does not transform the current batch payout through a reputation average. History affects assignment rather than adding directly to the score. This choice lowers incumbent protection but makes current sample variation more important. Separate extraction and judging reputations become more consequential under v1 because a miner can be strong at producing candidate records and weak at evaluating disputes, or the reverse.

7.5 Duplicate identities and marginal information

Social value depends on information rather than account count. Exact copies, shared unusual errors, highly overlapping source spans, and persistent lead-lag behavior can indicate one strategy spread across several identities. At the same time, independent methods can converge on the same correct answer.

The ecosystem addresses this problem with several tools:

- exact-duplicate and plagiarism penalties;
- first-contributor rules;
- similarity reductions;
- account binding and contribution eligibility;
- novelty or diversity bonuses;
- correlation and lead-lag analysis;
- marginal-contribution calculations.

Gittensor binds work to authenticated accounts and accepted repositories. CliqueAI reduces the value of exact copies while rewarding diverse strong solutions. Tag101 discounts semantic repetition. Hah incorporates uniqueness and marginal contribution relative to the crowd. Together, these mechanisms support a broad principle: redundant identities should not multiply reward automatically.

The prospective CLAIMS design applies that principle to semantic scientific production. A behavioral family can be estimated from candidate overlap, shared errors, source-span overlap, declared methods, and historical co-performance. One family rank and leave-family-out contribution would make reward less sensitive to identity splitting.

Classification error remains the central limitation. Independent convergence must not be mistaken for common ownership, and one adaptive strategy must not evade grouping through superficial variation. Family treatment is valuable only when its false merges and false splits are measured.

7.6 Memory, entry, and exploration

Short memory makes ranks noisy, while long memory can shelter incumbents and discourage entry. A workable market must exploit methods with demonstrated quality and continue to test new or under-observed methods.

CLAIMS currently uses qualification, performance, and rotation in assignment. Prospective family-level selection preserves the same exploration principle while reducing repeated assignment to redundant strategies. The relevant diversity measure is therefore the effective number of behaviorally distinct methods that receive evaluation and contribute accepted information, not the nominal number of registered UIDs.

Table 7 compares reward backbones, allocation rules, memory, and the treatment of repeated information.

Table 7. Reward and competition architectures.

Subnet	Backbone	Allocation	Memory	Treatment of information
CLAIMS v0	Verified contribution plus tournament	Fixed top-five rank schedule	Current batch; history affects assignment	Canonical items and exact-copy diagnostics
Apex	Benchmark frontier	Winner-take-all	Persistent current leader	Best valid complete solution
Data Universe	Contribution ledger	Proportional normalized value	Persistent credibility	Fresh, scarce, nonduplicative inventory
Gittensor	Contribution ledger	Proportional relative score	Rolling history and decay	Accepted unique code contributions
CliqueAI	Benchmark tournament	Proportional eligible performance	Long-memory reputation	Diversity discount for copies
cascade	Frontier ratchet	Published payout path incomplete	Persistent king and lineage	Statistically credible downstream improvement
Hah	Delayed skill market	Proportional after configured burn	Rolling and decayed	Edge, uniqueness, and marginal contribution

The table makes the CLAIMS hybrid visible. Its record is assembled like a contribution ledger, but its v0 payout is transformed like a tournament. Measuring how that combination affects method

diversity, entry, and accepted information is therefore part of the operating comparison rather than a separate design question.

8 Validators, Integrity, and Decentralization

Validation is a production process rather than a single act. A subnet can decentralize weight submission while keeping task generation, data access, model judgment, or canonicalization under one operator. The relevant comparison therefore follows the entire path from task construction to weights and asks which stages another validator can reproduce independently.

8.1 The functions inside validation

A complete validation stack can include:

1. task generation and input preparation;
2. assignment and commitment;
3. artifact transport and integrity checks;
4. deterministic validity and grounding;
5. hidden-test or holdout construction;
6. semantic or model-backed judgment;
7. consensus, abstention, or escalation;
8. canonicalization and record acceptance;
9. score construction and weight submission.

One validator can reproduce one stage while trusting another. Desarch validators can compare service outputs but depend on protected search and provider services. NIOME validators rely on private biological models and distributions. NOVA's model evaluation is reproducible conditional on target preparation and pinned checkpoints. Data Universe can revisit public sources while treating demand and desirability settings as governed inputs.

Decentralization therefore depends on which substantive functions are independently performed, not only on the number of parties submitting weights.

8.2 Independent replay

Independent replay is strongest when another validator can reconstruct the task, run the submission under the same conditions, reproduce the score, and verify the weight transformation. It is weaker when validators merely recompute arithmetic over records supplied by an operator.

Cryptographic proof can reduce the need to rerun a computation, but it moves trust to the circuit, the proof system, and the relation between the proved computation and the service users actually want. A proof of execution does not by itself establish semantic quality.

CLAIMS v0 preserves signed artifacts, source payloads, Bronze, candidate comparisons, adjudication decisions, Silver units, paper scores, exclusions, ranks, and proposed weights. Those records make technical replay possible. Substantive evaluation remains owner-operated, so an audit trail does not yet amount to independent production validation.

8.3 Centralized truth boundaries

Private holdouts, owner-run models, proprietary judges, customer task feeds, external interfaces, and reference datasets recur across the comparison set. Their presence does not by itself invalidate a mechanism. The material questions are whether the dependency is versioned, whether it can change silently, whether outputs can be replayed, whether evaluator error is measured, and whether another operator can reproduce the task-to-weight path.

CLAIMS has several such boundaries: paper acquisition, source parsing, Bronze construction, model-backed rigor checks, semantic adjudication, and prospective Gold. Release identifiers and stored artifacts make changes observable. Independent assurance still requires multiple evaluators that use genuinely distinct processes.

8.4 Origin blindness and evidence awareness

CLAIMS removes origin labels from semantic candidate packets. The judge does not receive the information needed to prefer Bronze because it is Bronze or to favor a known miner. This closes one direct route for identity and reference bias.

The judge does receive evidence records, source quotations, and source spans because the target is grounded scientific fidelity rather than fluent text. Origin blindness limits provenance shortcuts; evidence awareness limits acceptance of plausible but unsupported statements. A semantic judge without both controls can reward style, identity cues, or shared model priors while appearing to achieve high agreement.

8.5 Common-mode error

Several nominally independent validators can remain highly correlated when they use the same foundation model, prompts, parser, embeddings, reference set, and private service. Common-mode error is particularly serious when producers, challenge generators, and judges inherit the same model-family blind spot.

Assurance therefore comes from diversity of evaluation methods rather than validator count alone. Deterministic source checks, distinct model families, trusted human review, synthetic challenges, random re-audit, and later independent resolution provide different error signals. Prospective validator expansion improves CLAIMS only to the extent that new validators perform substantive independent work.

8.6 Current and prospective Claims validation

The current path and the proposed assurance layers should remain visually and analytically separate. Figure 7 shows which stages already determine v0 scores and which stages remain prospective.

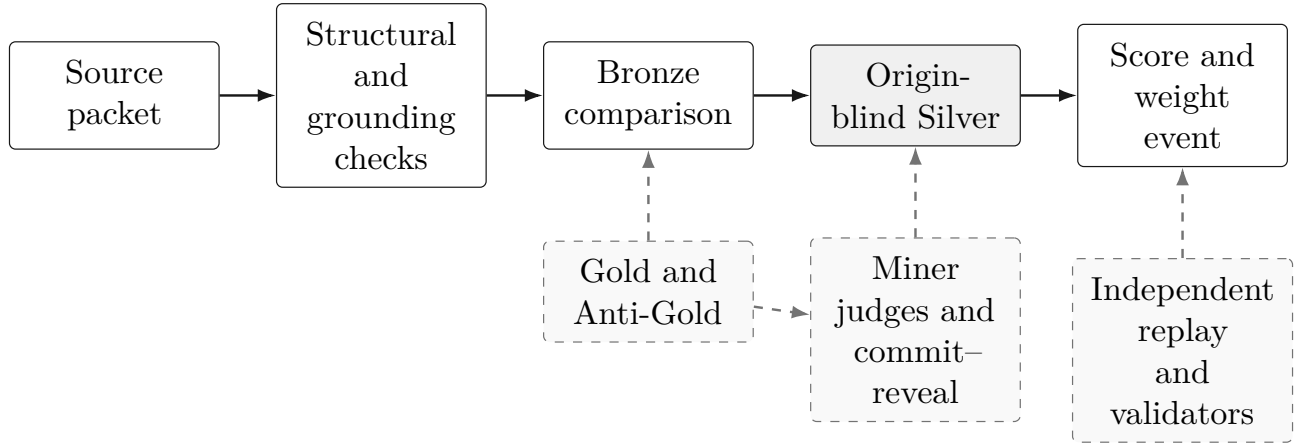


Figure 7. Current and prospective validation in Claims. Solid boxes form the v0 path; dashed boxes show proposed assurance layers whose value depends on independent information.

The dashed components add value only when they contribute independent information. More validators using the same models and source pipeline may improve availability without materially reducing semantic common-mode error.

8.7 Proof of assurance as a comparator

Provenance provides a useful boundary case because it verifies evidence that a workflow occurred, whereas CLAIMS must also judge the semantic object produced by that workflow.

Operating-status note. Provenance is recorded at pilot or testnet status rather than established production. The discussion refers to its published proof-of-assurance mechanism and reported data.

Provenance requires miners to return ordered evidence bundles containing traces, hashes, timestamps, nonce reflection, and outputs. Hard checks of integrity, task compliance, time, and completeness come before softer dimensions such as accuracy, reasoning quality, and efficiency. Unique nonces, hash chains, canary tasks, and replay detection constrain fabricated workflow evidence.

CLAIMS artifacts also preserve more than a final answer: source spans, evidence, trace, metadata, hashes, and signatures all remain available. The object being assured differs. Provenance asks whether a prescribed workflow was followed. CLAIMS asks whether a structured scientific representation is faithful and useful. A complete trace can prove that a process ran without proving that its semantic conclusion is correct.

Table 8 summarizes reported validator topology, reproducible layers, and retained dependencies.

Table 8. Validator topology and principal trust boundaries.

Subnet	Reported topology	Reproducible layer	Principal retained dependency
CLAIMS	One owner-operated substantive validator in v0; independent replay prospective	Artifacts, checks, comparisons, scores, and proposed weights	Corpus, Bronze, judges, and current validator operation
Data Universe	Independent replication	Public-source spot checks and scoring path	Source access and demand preferences
NOVA	Independent replication	Pinned model evaluation conditional on task assets	Target preparation and scientific models
NIOME	Authoritative service or trainer	Structured validation stages in the public core	Private biological models and reference distributions
cascade	Shared state with local recomputation	Generator commitments, seeds, and training comparison	Private series and authoritative payout path
Tag101	Authoritative service or trainer	Public scoring concepts	Post interface, cluster state, semantic models, and weight service
Minos	Published independent-replication design	Published hidden-genome rerun architecture	Current generator, evaluator, metric, and weight path not independently verified
ByteLeap	Authoritative service or trainer	Sampled proof verification	Lease marketplace, routing, and resource state

The topology comparison separates auditability from substantive independence. Preserved artifacts are necessary for replay, but assurance rises only when another evaluator can reproduce the scientific judgment rather than merely confirm the arithmetic.

9 Closest Comparisons on a Common Template

The preceding sections compare subnets by theme. A fixed template makes the nearest relationships easier to compare directly: product, miner task, validator task, truth source, persistent output, reward structure, integrity controls, principal commonality, and decisive difference. The exercise does not produce a single ranking; it shows which dimension carries each resemblance.

9.1 Claims and Minos

Minos remains the closest composite mechanism relative because the two systems combine scientific work, private production methods, hidden evaluation, and outputs that can improve later production. Table 9 shows that the answer key, rather than the broad workflow, creates the decisive separation.

Table 9. Claims and Minos on common comparison dimensions.

Dimension	CLAIMS	Minos
Product	Canonical scientific claim–evidence records	Variant-calling performance and synthetic genomic outputs
Miner task	Produce source-grounded structured records	Configure or develop a variant-calling pipeline
Validator task	Check, compare, adjudicate, canonicalize, and score	Rerun on hidden synthetic genomes and compare with labels
Truth source	Direct source checks plus semantic Silver; prospective Gold and Anti-Gold	Hidden inserted mutations and private benchmark truth
Persistent output	Evidence graph, difficult cases, and improved references	Validated synthetic genomes and improved calling methods
Reward logic	Absolute quality followed by concentrated rank	Published relative-performance contest; exact allocation not established
Principal commonality	Scientific hidden evaluation over private methods	Scientific hidden evaluation over private methods
Decisive difference	Partial semantic answer key	More complete synthetic answer key

Minos shows that Bittensor already supports the broad combination of standardized scientific work, hidden tasks, validator reruns, and productive benchmark output. CLAIMS bears an additional institutional burden because its benchmark cannot fully identify the desired record.

9.2 Claims and Data Universe

Both systems build persistent, source-grounded data assets. Data Universe maintains fresh inventories; CLAIMS maintains canonical scientific representations. Data Universe can use spot checks to update a miner’s credibility across an inventory. CLAIMS cannot infer the truth of one item solely from the miner’s broader credibility, so verification state remains attached to each canonical item.

The systems can occupy successive stages in one production stack. A generalized acquisition layer can supply documents and metadata, while CLAIMS adds semantic extraction, evidence relations, and versioned correction. They substitute mainly when customers prefer raw or lightly normalized source access to curated scientific propositions.

9.3 Claims and Orion

Both systems issue structured assignments, receive signed artifacts, impose hard validity gates, use hidden replay, and reward accepted production. Orion’s campaign architecture is therefore a close operational analog.

The final CLAIMS object persists across campaigns and must reconcile competing records. Source spans, claim identity, evidence links, semantic equivalence, and supersession form part of the commodity itself. Orion demonstrates campaign-based distributed data production; it does not by itself solve the maintenance of canonical scientific knowledge.

9.4 Claims and ReadyAI

ReadyAI and CLAIMS both depend on semantic extraction and an external evaluator. Their shared risk is evaluator centrality: reference records, embeddings, and language-model judges can define the task more strongly than the source material.

CLAIMS responds by exposing source packets, evidence, origin blindness, and canonicalization. The accepted record must do more than resemble a reference label. It must remain auditable against the paper and preserve distinctions that a semantic similarity score can collapse.

9.5 Claims and Tag101

Tag101 provides the clearest comparison for semantic novelty and repetition. It reduces the value of reused tag sets and rewards fit with current content clusters. CLAIMS faces the same redundancy problem at a more structured semantic level.

A simple repetition penalty is insufficient for scientific records. Independent corroboration has evidentiary value, and one valid claim can have several acceptable formulations. Canonicalization first identifies the shared information; prospective behavioral-family and marginal-contribution rules then estimate how much additional value each method supplied.

9.6 Claims and cascade

Both systems can produce a baseline that improves through past accepted work. cascade promotes a generator after common downstream tests. CLAIMS promotes reference and judge releases after corrected examples and external evaluation. In each case, the benchmark becomes endogenous to earlier winners.

That feedback creates a common governance requirement. Releases must be frozen, evaluation must remain held out, and historical results must remain interpretable under the rule that produced them. cascade can define improvement through one maintained forecasting environment; CLAIMS must protect a broader evidence object intended for several downstream uses.

9.7 Claims and Desearch or rec4ll

Desearch is an active query service, whereas the current rec4ll mechanism is unresolved. Both illustrate the user-facing alternative to pre-curated evidence: retrieve current material when a question arrives rather than maintain a complete graph in advance.

CLAIMS becomes more distinct as use becomes repeated, cross-paper, and machine-integrated. One claim–evidence record can support retrieval, contradiction detection, training, and later correction. Retrieval becomes complementary when it searches over this structured layer rather than reconstructing it for every request.

9.8 Claims and Apex or CliqueAI

Apex and CliqueAI show how concentrated competition works when validity is inexpensive to establish. Their evaluators can reject invalid output and compare eligible solutions on a common objective. CLAIMS uses similar rank intensity but cannot rely on one complete submission or one scalar truth source.

This comparison exposes a design tension already present in v0. Concentrated rewards can induce frontier effort, while partial verification and batch sampling can make ranks noisier. Operating evidence must therefore show whether payout concentration preserves useful method diversity and entry while still rewarding clear improvement over Bronze.

Table 10 condenses the pairwise comparisons into the combination of features that distinguishes CLAIMS.

Table 10. The combination of product and mechanism features has no exact twin in the featured comparison set.

Dimension	Claims	Minos	Data Uni-verse	Orion	ReadyAI	cascade
Scientific evidence product	Yes	Genomic inference	No	No	No	No
Persistent external asset	Yes	Yes	Yes	Yes	Recurring records	Champion generator
Source-span grounding	Core	Sequencing reads	Source re-visit	Assignment dependent	Passage dependent	No
Complete cheap truth	No	Selected synthetic cases	Partial spot checks	Hidden samples	Reference + judge	Downstream benchmark
Semantic adjudication	Core	Limited	Limited	Judge dependent	Core	No
Synthetic challenge potential	Prospective broad role	Core hidden genome	Limited	Hidden samples	Possible	Private holdout
Canonical item identity	Core	Not central	Duplicate control	Duplicate control	Unique labels	Incumbent comparison
Reversible lineage	Intended graph property	Dataset lineage	Inventory history	Submission history	Score history	Champion lineage
Marginal-information objective	Prospective	Not established	Scarcity discount	Useful accepted output	Distinct labels	Credible improvement
Current operating status	Active v0	Published current design	Active	Active	Active	Active

No single row is unique to CLAIMS. The absence of an exact twin comes from combining a persistent scientific evidence product, partial semantic truth, canonical item identity, reversible lineage, and a proposed marginal-information objective in one system.

10 Operating and Market Context

Performance has three meanings that must remain separate: production performance, product-use performance, and token-market performance. A subnet can have strong token performance without established product use, or strong technical operation without a large market value. The comparison reports each class without treating one as evidence for another.

10.1 Production performance

Production performance asks whether the mechanism completes its assigned task reliably. For CLAIMS, the relevant measures include:

- valid artifact completion and source-grounding success;
- paper and claim throughput;
- Bronze, Silver, and final precision and recall on trusted audits;
- accuracy of claim–evidence relations;
- net correction of Bronze after subtracting damage to correct content;
- score dispersion and rank stability;
- validator exclusions, failure stages, and replay agreement;
- cost and latency by pipeline stage;
- the effective number of behaviorally distinct miner methods;
- re-audit reversals and reference-release regressions.

These measures answer different failure modes. High throughput cannot compensate for a failed quality floor. A large nominal miner population cannot compensate for one method supplying nearly all accepted information. A high challenge score cannot substitute for evidence that challenge performance transports to ordinary records.

10.2 Product-use performance

Product-use performance asks whether records are consumed outside the scoring loop. Observable measures include delivered datasets, interface use, repeated queries, customer-specific schemas, downstream integrations, corrections submitted by users, and reuse of one record across applications. These measures belong to the data product rather than the token market.

The available comparison data do not establish one commercial measure across the featured subnets. Some systems meter services, some run research tournaments, and some build assets whose use does not enter the reward path directly. Service revenue, product use, and emissions therefore remain separate rather than being combined into one performance number.

10.3 Operating maturity and earning activity

The dated snapshot separates the state of a mechanism from economic participation in it. *Active* means that a current mechanism path is identified. *Published design* means that the mechanism is specific enough to analyze although the complete production path is not independently established. *Pilot/testnet* indicates limited deployment. *Unresolved* means that no definite current contest or allocation path can be attributed.

Earning-miner activity is classified separately as broad, active, thin, or none. The labels can diverge. Orion, Cacheon, and ByteLeap are recorded as active even though no earning miners appear in the snapshot. Minos has reported active earning-miner activity while its implementation remains classified as a published current design with limited verification. Both facts matter, and neither replaces the other.

10.4 Dated token-market snapshot

The market observations in Table 11 are dated 25 August 2026. Alpha return is measured against TAO. Pool TAO change is gross reserve change rather than identified net capital inflow. Historical observations apply only to the current registration instance. *Too new* means that the instance did not exist at the comparison date.

Table 11. Selected subnet token-market context as of 25 August 2026.

Subnet	Registered	Market value	Rank	6m return	12m return	Pool TAO	Alpha staked	Operating status
SN111 Claims	22 May 2025	25,523 TAO	42	+70.3%	+346.0%	6,677	62.4%	Active v0
SN107 Minos	02 Feb 2026	95,756 TAO	12	+531.5%	<i>Too new</i>	22,146	60.9%	Published design
SN68 NOVA	21 Feb 2025	109,346 TAO	11	+78.8%	+64.5%	44,268	52.4%	Active
SN55 NIOME	14 Jan 2025	14,435 TAO	74	−26.1%	+34.5%	6,769	50.2%	Active
SN1 Apex	13 Apr 2023	38,198 TAO	27	−32.2%	−45.7%	25,007	29.5%	Active
SN83 CliqueAI	29 Mar 2025	33,987 TAO	31	+158.2%	+505.9%	8,862	67.2%	Active
SN13 Data Universe	11 Dec 2023	31,850 TAO	34	−33.3%	−59.2%	20,734	31.3%	Active
SN27 Orion	16 Nov 2023	11,421 TAO	90	−37.7%	+13.4%	6,360	35.2%	Active
SN33 ReadyAI	12 May 2024	23,555 TAO	49	−51.9%	−68.9%	15,112	32.0%	Active
SN101 Tag101	02 May 2025	11,162 TAO	91	−44.8%	−100.0%	3,990	47.9%	Active
SN91 cascade	27 Feb 2026	8,672 TAO	98	<i>Too new</i>	<i>Too new</i>	3,544	43.6%	Active
SN22 De-search	26 Dec 2023	15,309 TAO	68	−35.4%	−3.4%	7,812	39.2%	Active
SN31 rec4ll	25 Dec 2025	7,105 TAO	100	+12.8%	<i>Too new</i>	2,034	47.6%	Unresolved
SN74 Git-tensor	08 Mar 2025	16,118 TAO	67	−53.1%	+161.1%	6,223	55.1%	Active
SN128 Byte-Leap	25 Jun 2025	12,257 TAO	86	−22.8%	+14.6%	4,109	66.5%	Active

These figures describe scale, staking, and relative token performance. They do not establish scientific accuracy, technical quality, customer demand, commercial revenue, or sustainable operation.

The CLAIMS market profile lies between several comparison groups. Its market value is below the large scientific-design and genomics comparators but above several data and mechanism analogs. Its dated six- and twelve-month alpha returns are strong relative to many selected peers. Those observations describe token-market history rather than the causal effect of the CLAIMS product or mechanism. Registration age, liquidity growth, alpha supply dynamics, changing emissions, and market attention all vary over the same period.

10.5 Neutral operating metrics for Claims

Useful operating measures should correspond to specific claims about the mechanism. Table 12 pairs each measure with the companion information needed to interpret it.

Table 12. Operating measures and the mechanism claims they test.

Measure	Interpretation	Required companion information
Trusted-audit precision and recall	Fidelity of claims and evidence links	Audit composition, error classes, confidence intervals, and Gold provenance
Net lift over Bronze	Corrected Bronze errors minus damage to correct Bronze content	Discovery rate, false challenges, and Silver error by case type
Verified data velocity	Audited records or claim–evidence links produced per unit time	Raw throughput, latency, verification-state mix, and reversal rate
Effective miner-family count	Number of independent methods supplying accepted information	Cluster definitions, reward shares, and false-merge and false-split tests
Split-gain ratio	Whether identity duplication raises combined reward	Assignment, evidentiary weight, and payout under simulated splits
Validator assurance	Ability to detect seeded defects without harming clean controls	Recall, false-positive rate, abstention, calibration, and replay agreement
Cost per verified record	Operating efficiency of Bronze, Silver, Gold, storage, and audit	Quality floor, miner break-even, validator cost, and domain mix
External record use	Reuse beyond the scoring loop	Interface calls, datasets, repeated users, integrations, and corrections

The measures are not interchangeable. A fast system can be wrong, a highly precise system can omit central claims, a large registered population can contain one effective method, and a low cost

per record can reflect a weak verification standard. Quality, velocity, diversity, assurance, cost, and external use must therefore remain separate through the conclusion.

11 Conclusion

CLAIMS sits between two established parts of the Bittensor ecosystem. Its product resembles persistent data construction and semantic extraction: miners turn source material into reusable records. Its mechanism resembles scientific benchmarks, contribution ledgers, frontier contests, and hidden-test markets: validators must make useful work observable and decide how much independent information it adds.

Minos draws the boundary most clearly. Both systems combine a scientific task, competition over private methods, hidden evaluation, and output that can improve future production. Hidden synthetic genomes, however, provide a more complete scalar answer key than scientific extraction. A scientific record can fail through omission, semantic equivalence, lost qualification, incorrect evidence, weak source grounding, or genuine ambiguity. Those failure modes make selective judgment and unresolved states part of the mechanism rather than exceptions to it.

The comparison therefore rests on three separations. Product proximity is distinct from mechanism proximity. Fidelity to a paper is distinct from the later truth of the paper’s scientific conclusion. Measuring whether a producer works reliably is distinct from establishing or correcting the canonical content of a real case. Bronze, Silver, Gold, Anti-Gold, item-level lineage, and evaluator testing assign different roles to those problems instead of hiding them inside one score.

The active and prospective systems remain separate as well. V0 implements structured extraction, deterministic checks, Bronze comparison, origin-blind Silver adjudication, multiplicative paper scoring, and concentrated rank rewards. Gold, Anti-Gold, post-commit sampling, miner judging, behavioral-family treatment, marginal contribution, stronger independent validation, and canonical production ingestion remain proposed extensions whose value depends on validity, transport, concealment, scoreability, coverage, and replay.

CLAIMS is therefore neither a simple benchmark market nor a generic data market. It is an effort to build a reusable scientific evidence asset while keeping accepted improvements, participant rewards, and every layer of trust inspectable when complete ground truth is unavailable.

A Relevant-Subnet Comparison Matrix

This appendix records the full featured comparison set on one template. *Active* indicates an identified current mechanism path. *Published design* indicates a specific current mechanism whose complete production path has not been independently verified. *Pilot/testnet* indicates limited deployment, and *unresolved* indicates that the current mechanism cannot be attributed with confidence.

Table 13. Featured subnet comparison matrix.

SN	Name	Product	Backbone	Truth source	Integrity emphasis	Observability	Status
1	Apex	Optimization solutions	Benchmark tournament	Deterministic benchmark	Hidden cases, sandbox, replay	Public reproducible path	Active
13	Data Universe	Public-web inventories	Contribution ledger	Public-source revisit	Duplication, credibility, consensus	Public reproducible path	Active
14	Cacheon	Compute optimizations	Frontier ratchet	Reproduced benchmark	Build and quality gates	Public reproducible path	Active
22	Desearch	Search and research responses	Metered service	Provider checks and service judgment	Authentication, validity, corroboration	Public reproducible path	Active
27	Orion	Structured training data	Contribution ledger	Hidden samples and judges	Signatures, replay, duplicate gates	Public core with dependency	Active
31	rec4ll	Retrieval and persistent memory	Unresolved	No established current truth source	No established current controls	Docs only	Unresolved
33	ReadyAI	Semantic labels and tags	Contribution ledger	Reference records and AI judgment	Uniqueness and qualification gates	Public core with dependency	Active
34	BitMind	Synthetic-media detection and generation	Benchmark tournament	Known media labels and private holdouts	Fresh data, standardized execution	Public core with dependency	Active
55	NIOME	Synthetic CRISPR data	Contribution ledger	Private biological models and distributions	Staged validity and plausibility	Public core with dependency	Active
57	Hah	Sports forecasts	Delayed skill market	Future outcomes and market lines	History, uniqueness, marginal contribution	Public reproducible path	Active
68	NOVA	Molecules and nanobodies	Benchmark tournament	Pinned predictive models	Commit-reveal, duplicates, deterministic seeds	Public reproducible path	Active
74	Gittensor	Open-source contributions	Contribution ledger	Public merged work and replay	Account binding, duplicate controls	Public reproducible path	Active
83	CliqueAI	Maximum-clique solutions	Benchmark tournament	Formal graph validity	One-use tasks, duplicate and diversity rules	Public reproducible path	Active
84	ChipForge	Hardware designs	Frontier ratchet	Hidden functional tests and EDA	Hashes, signed requests, rate limits	Public core with dependency	Active
87	Provenonce	Workflow assurance bundles	Contribution ledger	Formal integrity plus judged quality	Nonces, hash chains, replay detection	Public core with dependency	Pilot

SN	Name	Product	Backbone	Truth source	Integrity emphasis	Observability	Status
91	cascade	Synthetic forecasting data	Frontier ratchet	Private downstream test series	Immutable artifacts, future-block seeds, replay	Public reproducible path	Active
101	Tag101	Semantic tags for social posts	Contribution ledger	AI judgment and private cluster state	Repetition and novelty controls	Public core with dependency	Active
107	Minos	Variant-calling competition	Benchmark tournament	Hidden synthetic genomes	Fresh hidden tasks and replay	Published design; limited verification	Published design
111	Claims	Scientific claim-evidence records	Contribution plus rank tournament	Source checks, Silver; prospective Gold	Signatures, grounding, origin blindness, replay	Public core with dependency	Active v0
128	ByteLeap	Compute rental and idle proofs	Metered service	Usage records and sampled proofs	Commit before sampling, quotas, proofs	Public core with dependency	Active

The matrix should be read across rows rather than as a ranking. Product, truth source, integrity controls, observability, and status answer different questions and can point in different directions for the same subnet.

B Comparator Profiles

The profiles below preserve the status, product, mechanism, and CLAIMS comparison for the nearest product or mechanism relatives without turning each short profile into a separate appendix section.

Minos.

Status: Published current design with reported active earning-miner activity; complete implementation path not independently verified. **Product:** Variant-calling methods and validated synthetic genomic tasks. **Mechanism:** Hidden synthetic genomes, validator rerun, false-positive and missed-variant penalties, and relative-performance competition. **Claims comparison:** The closest composite scientific mechanism, with a stronger synthetic answer key and less semantic ambiguity.

Data Universe.

Status: Active. **Product:** Fresh, useful, retrievable public-web inventories. **Mechanism:** Contribution value combines volume, freshness, demand, duplication, and miner credibility inferred from spot checks. **Claims comparison:** A similar persistent asset with source retrievability, but less semantic canonicalization.

Orion.

Status: Active, with no earning miners in the dated snapshot. **Product:** Campaign-defined structured training datasets. **Mechanism:** Signed assignments, schema and availability gates, hidden replay, and accepted contribution across quality and useful volume. **Claims comparison:** A close operational workflow; CLAIMS adds scientific identity and lineage that persist across campaigns.

ReadyAI.

Status: Active. **Product:** Semantic labels, tags, and mappings extracted from text. **Mechanism:** Reference comparison, semantic similarity, coverage, correctness, and uniqueness. **Claims comparison:** A close semantic-extraction relative; CLAIMS adds claim-evidence relations and source-grounded canonicalization.

Tag101.

Status: Active. **Product:** Timely semantic tags for social posts. **Mechanism:** Cluster fit, post value, novelty, averaging, and repetition history. **Claims comparison:** A strong analog for semantic redundancy; CLAIMS must also preserve independent corroboration.

cascade.

Status: Active. **Product:** Synthetic training data that improve a common forecasting model. **Mechanism:** Immutable generators, future-block seeds, common training budgets, inexpensive heats, a bootstrap challenger gate, and a persistent king. **Claims comparison:** A strong analog for an endogenously improving baseline and recursive contamination risk.

Desearch.

Status: Active. **Product:** Query-time search, retrieval, and synthesis. **Mechanism:** Service quality, validity, corroboration, latency, and moving miner scores. **Claims comparison:** An interface substitute and a potential distribution complement.

rec4ll.

Status: Unresolved. **Product:** Announced retrieval and persistent-memory service. **Mechanism:** The current benchmark and reward path are not established. **Claims comparison:** Product-level architecture only; no reliable mechanism comparison.

Apex and CliqueAI.

Status: Active. **Product:** Objectively verifiable optimization solutions. **Mechanism:** Hard validity gates, common benchmarks, concentrated or diversity-adjusted rewards, and independent replay. **Claims comparison:** Boundary cases showing why aggressive frontier rewards are easiest to support when truth is inexpensive.

Together, the profiles show that the nearest product relative and the nearest mechanism relative need not be the same subnet. That distinction is the organizing principle of the main text.

C Claims Current–Prospective Crosswalk

The crosswalk keeps the active v0 mechanism distinct from proposed directions at the level of each object, role, reward rule, and production dependency.

Table 14. Current Claims mechanism and prospective directions.

Object	Current v0	Prospective direction
Bronze	Versioned reference extraction used in live comparison	Frozen release promoted through held-out Gold and regression gates
Silver	Origin-blind machine adjudication and canonical units	Gold-calibrated, confidence-aware adjudication with escalation
Gold	Not incentive-bearing	Trusted real cases for quality estimation, resolution, discovery, and release gates
Anti-Gold	No active reward lane	Hidden controlled mutations for reliability, effort, and training
Miner role	Structured extraction	Separate extraction and blinded judging
Evaluation scope	Returned scoring papers follow the full Silver path	Full commitment followed by representative post-submit sampling
Identity	UID and exact-copy observations	Behavioral families, correlated-strategy controls, one family rank
Value attribution	Agreement with final Silver and accepted improvement	Leave-family-out marginal information and bounded corroboration
Reward	Top-five rank schedule	Family-level ranking with separate extractor and judge treatment

Object	Current v0	Prospective direction
Database	Calibration artifacts; no automatic production ingestion	Verification-tiered canonical records with reversible lineage
Validator topology	One owner-operated substantive validator	Independent replay and multiple substantive evaluation processes
Release control	Versioned operational components	Frozen full-stack releases gated on external holdouts and audits

The prospective column is not a description of current operation. Each addition becomes part of the incentive-bearing mechanism only after its measurement, failure conditions, and replay path are specified.

D Detailed Market Table

Table 15 expands the dated market snapshot with pool and staking changes. The same cautions as in the main text apply: the observations describe token-market history as of 25 August 2026 and do not establish product quality, scientific accuracy, customer demand, revenue, or causal effects of mechanism design.

Table 15. Selected token-market measures and changes as of 25 August 2026.

Subnet	Market value	6m return	12m return	Pool TAO	6m pool	12m pool	Staked	6m staked	12m staked
Claims	25,523	+70.3%	+346.0%	6,677	+2,575	+5,930	62.4%	+10.1 pp	+18.5 pp
Minos	95,756	+531.5%	<i>Too new</i>	22,146	+22,105	<i>Too new</i>	60.9%	+3.6 pp	<i>Too new</i>
NOVA	109,346	+78.8%	+64.5%	44,268	+16,398	+26,668	52.4%	+14.3 pp	+11.1 pp
NIOME	14,435	−26.1%	+34.5%	6,769	−92	+3,870	50.2%	−3.7 pp	+13.9 pp
Apex	38,198	−32.2%	−45.7%	25,007	−3,734	+4,661	29.5%	−3.3 pp	−5.8 pp
CliqueAI	33,987	+158.2%	+505.9%	8,862	+5,122	+8,862	67.2%	+18.8 pp	−31.4 pp
Data Universe	31,850	−33.3%	−59.2%	20,734	−3,711	+666	31.3%	−4.0 pp	−10.3 pp
Orion	11,421	−37.7%	+13.4%	6,360	−884	+3,344	35.2%	−8.4 pp	−5.1 pp
ReadyAI	23,555	−51.9%	−68.9%	15,112	−5,484	−888	32.0%	−10.7 pp	−23.2 pp
Tag101	11,162	−44.8%	−100.0%	3,990	+375	+3,986	47.9%	−2.5 pp	+6.9 pp
cascade	8,672	<i>Too new</i>	<i>Too new</i>	3,544	<i>Too new</i>	<i>Too new</i>	43.6%	<i>Too new</i>	<i>Too new</i>
Desearch	15,309	−35.4%	−3.4%	7,812	−1,177	+3,845	39.2%	−3.3 pp	−10.2 pp
rec4ll	7,105	+12.8%	<i>Too new</i>	2,034	+1,694	<i>Too new</i>	47.6%	−2.4 pp	<i>Too new</i>
Gittensor	16,118	−53.1%	+161.1%	6,223	−1,910	+4,876	55.1%	−4.3 pp	+4.5 pp
ByteLeap	12,257	−22.8%	+14.6%	4,109	+250	+3,771	66.5%	+10.2 pp	−1.7 pp

The changes are reported for the current registration instances. *Too new* indicates that the relevant comparison period did not exist, and percentage-point changes in staking should not be read as returns.

E Glossary

The glossary collects the paper-specific terms needed to interpret the comparison, especially the distinction between the current record, participant reliability, and prospective assurance.

Anti-Gold.

A hidden synthetic known-answer task created by a controlled transformation of a represented error class.

Behavioral family.

A group of identities whose outputs indicate one highly correlated information-producing strategy.

Bronze.

The economical reference extraction used as a common baseline.

Canonicalization.

The mapping of equivalent expressions to stable content units while preserving material differences and ambiguity.

Contribution ledger.

A mechanism that accumulates verified work rather than paying only for one current rank.

Frontier ratchet.

A mechanism in which an incumbent contribution persists until a qualifying challenger displaces it.

Gold.

Trusted information from real or known-answer cases used for resolution, calibration, discovery, or release gates.

Latent error class.

A failure mode that the system cannot yet generate, conceal, score, and transport as a synthetic challenge.

Marginal contribution.

The value that disappears when a miner or behavioral family is removed from the accepted information set.

Origin blindness.

The removal of candidate provenance that would let a judge favor a reference or miner identity directly.

Post-commit sampling.

Selection of the audited subset only after the participant commits to the full assigned work.

Represented error class.

A failure mode for which the platform has a valid, concealed, scoreable, and production-relevant challenge.

Silver.

The stronger machine-adjudicated and canonicalized record.

Truth production.

Establishing or correcting the canonical content of a real case, as distinct from estimating producer reliability.

Validator replay.

Reconstructing the task, evaluation, score, and weight path from preserved artifacts and fixed rules.

These definitions preserve the boundaries used throughout the paper. In particular, no verification tier or participant score is treated as a substitute for the scientific record itself.