Minos
Decentralized Infrastructure for DNA Mutation Inference & Benchmarking

Minos Team

February 2026

Abstract

As DNA sequencing costs have fallen from $1 million to under $200 per genome, the bottleneck in genomics has shifted from data generation to data analysis. **Variant calling**, the process of identifying DNA mutations from sequencing data, is the critical inference layer, and despite major investment from Google, the Broad Institute, and government agencies, no single tool is consistently the most accurate. Minos is a Bittensor subnet that addresses this through decentralized, incentive-driven optimization. Every 72 minutes, the platform generates a fresh challenge genome containing hidden synthetic mutations injected at the read level. Miners optimize variant calling tool configurations, and in later phases, submit entirely custom algorithms, while validators execute them in isolated containers and score the results against known ground truth. The best-performing miner receives all the emissions.

This continuous benchmarking produces three valuable outcomes. First, as a byproduct, Minos builds the **world's largest validated synthetic genome database**: over thousands of validated genomes in only the first year, each with confirmed ground-truth mutations. Second, by aggregating the top-performing models across tools and genomic contexts, Minos trains a **consensus variant caller** expected to exceed the accuracy of any individual method. Third, this consensus model will be deployed as **AI-powered variant calling infrastructure** where hospitals, biobanks, and pharmaceutical companies can submit sequencing samples and receive high-accuracy variant calls, reducing missing diagnoses and minimizing clinical errors through continuously validated, decentralized AI algorithms developed by Minos. This document provides a unified overview of the Minos architecture, incentive mechanism, scoring, security properties, and the road ahead.

1 The Genomics Data Explosion and Its Bottleneck

When the Human Genome Project completed in 2003, sequencing a single human genome cost roughly **$1 million** and took years. Laboratories could afford to sequence only a handful of samples per year. Today, that same genome can be sequenced for **$100–200** in under 24 hours. This 5,000-fold cost reduction has transformed genomics from a boutique research activity into an industrial-scale data operation: hospitals, biobanks, and direct-to-consumer companies now generate **thousands to potentially millions** of whole genomes per year.

But cheaper sequencing has shifted the bottleneck. Generating raw DNA data is no longer the hard part, *accurately analyzing* that data is. Every genome must be processed by specialized software that reads through billions of short DNA fragments and determines where an individual's DNA differs from the reference. This computational step, known as **variant calling**, is the inference layer of genomics. As sequencing throughput scales exponentially, the accuracy, speed, and reliability of this inference layer become the critical constraint on the entire field.

The genomics market reflects this shift, projected to more than quadruple from $37.95 billion in 2024 to $176.28 billion by 2034 (Figure 1). The demand is no longer for more sequencing machines, it is for better, faster, and more reliable analysis of the data those machines produce.

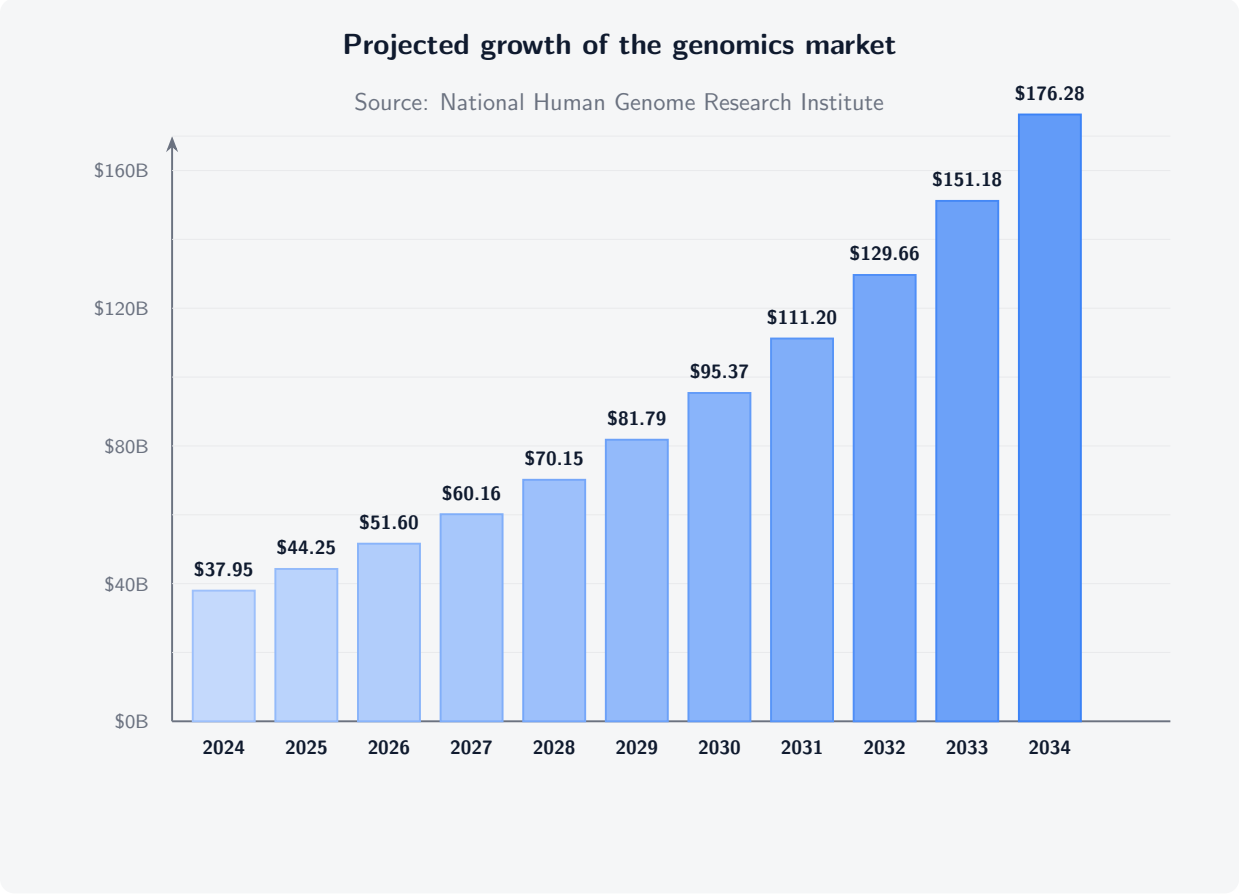

Figure 1: Projected growth of the global genomics market from $37.95 billion (2024) to $176.28 billion (2034). As sequencing becomes cheaper and clinical genomics adoption expands, the demand for accurate, scalable variant calling infrastructure grows proportionally. Source: National Human Genome Research Institute.

Every human genome contains roughly 4–5 million places where an individual's DNA differs from the reference human genome. Finding these differences, called **variant calling**, is the essential first step in nearly all genomic analysis: diagnosing genetic diseases, predicting drug responses, analyzing ancestry, and conducting large-scale biobank research.

Variant calling is essentially spell-checking a 3-billion-letter document. The software reads through the sequencing data and flags every position where the letters differ from the reference. Missing a real mutation (a false negative) or flagging a normal position as mutated (a false positive) both have consequences, especially in clinical settings.

The biggest names in technology and science recognize this bottleneck and are actively investing in it. **Google** developed DeepVariant, a deep-learning variant caller that won the FDA's precisionFDA Truth Challenge. The **Broad Institute**, a joint Harvard/MIT research center funded by the NIH and one of the most influential genomics institutions in the world, maintains GATK (Genome Analysis Toolkit), the most widely used variant calling framework in clinical and research settings. Government agencies including the **National Institute of Standards and Technology (NIST)** fund the Genome in a Bottle consortium, which produces the gold-standard truth sets that the entire field relies on for benchmarking. Meanwhile, companies like **Illumina** (the dominant sequencing hardware manufacturer) and **PacBio** are building their own variant calling pipelines to pair with their machines.

Yet despite this investment from trillion-dollar companies, federally funded institutions, and dedicated research consortia, the variant calling problem remains unsolved. Each of these players develops tools in isolation, benchmarks against static datasets, and publishes results that are difficult to compare across methods. There is no continuous, independent, head-to-head evaluation system, which is precisely the gap Minos fills.

1.1 From Sample to Genotype

Figure 2 shows the general process of genomic variant calling: a biological sample is collected, DNA is extracted, the DNA is sequenced into raw reads, and then specialized software tools process those reads to identify genetic variants (genotypes). Critically, different tools can produce different results from the same raw data.

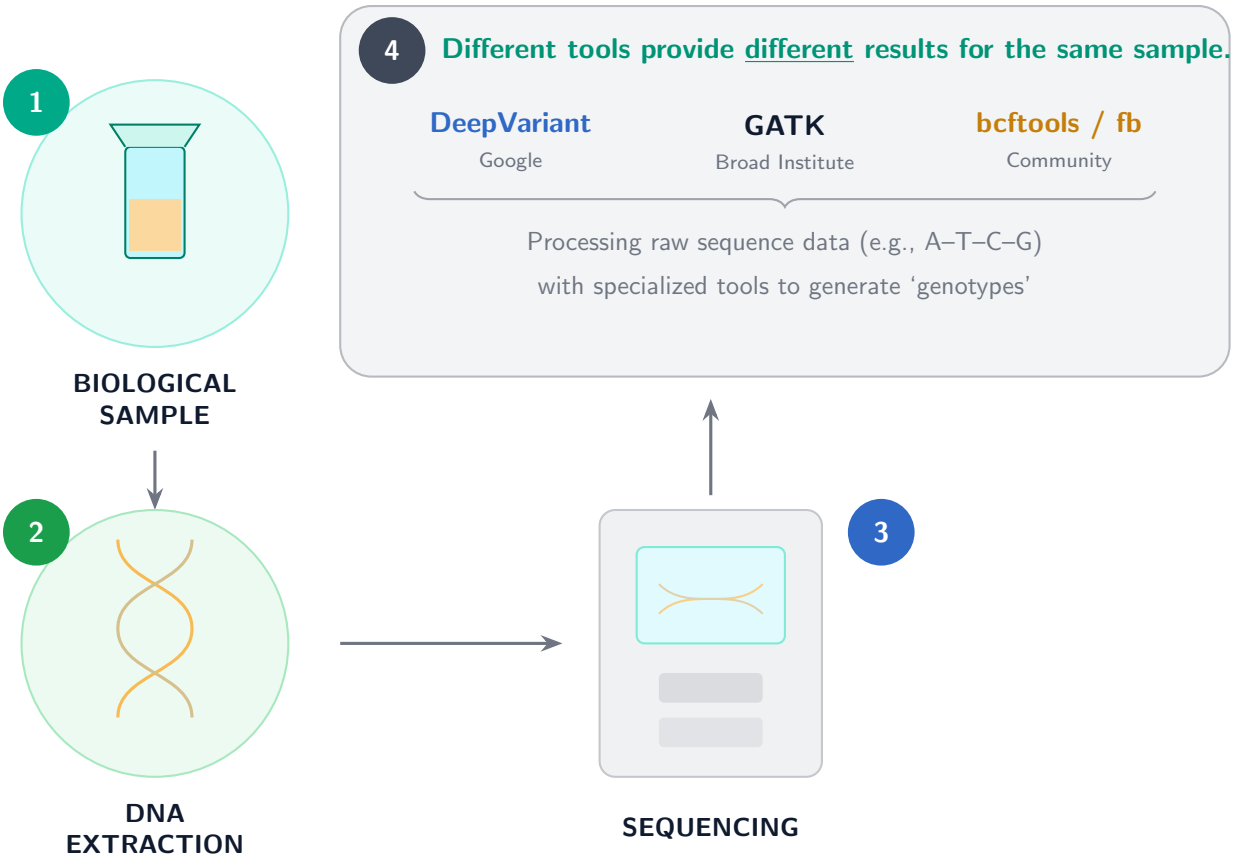

Figure 2: The genomics pipeline. A biological sample (Step 1) undergoes DNA extraction (Step 2) and sequencing (Step 3) to produce raw reads. Specialized variant calling tools (Step 4) then process these reads to identify genetic variants. Different tools can produce different genotype calls from the same data.

1.2 The Disagreement Problem

Figure 3 illustrates the core challenge: when multiple tools analyze the same samples, they often disagree on the correct genotype. No single tool is always right. The correct call (rightmost column) varies, and each tool has different strengths and blind spots.

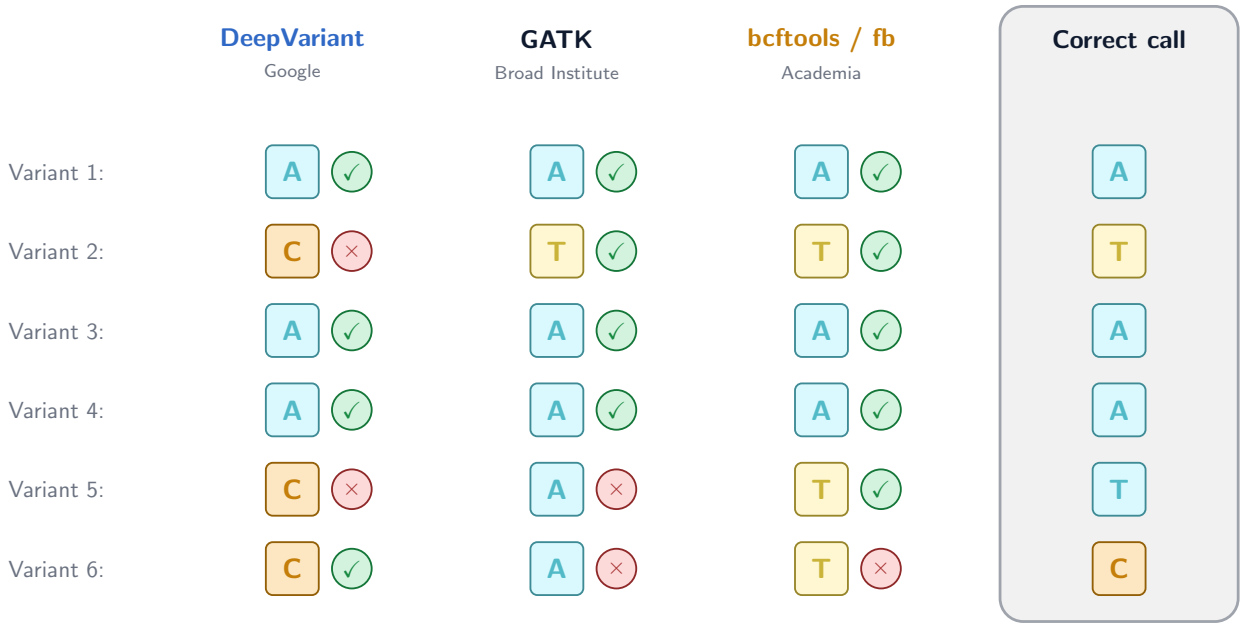

Figure 3: The disagreement problem in variant calling. Three different tools analyze the same six genomic positions and produce different genotype calls. No single tool is always correct: DeepVariant errs on Samples 2 and 5; GATK errs on Samples 5 and 6; bcftools errs on Sample 6. The correct call (right column) varies by position. This inconsistency is the fundamental problem Minos addresses through continuous, competitive benchmarking.

The problem today:

- **No continuous benchmarking**: Gold-standard truth sets exist (GIAB), but labs benchmark their own pipelines with no independent, ongoing verification.
- **Duplicated effort**: Thousands of organizations independently run and evaluate the same variant calling tools, wasting compute and human resources.
- **Stagnant innovation**: Labs stick with legacy pipelines because updating requires re-benchmarking everything from scratch.
- **Black-box vendor outputs**: Sequencing providers report accuracy metrics without independent verification.

What Minos does: Minos replaces this fragmented landscape with a single, decentralized, continuously operating benchmarking system. Miners compete to find the most accurate variant calling configurations, validators independently verify every result, and the best performer earns all the emissions.

2 How Minos Works

The approach is staged. In its current phase, Minos focuses on **hyperparameter optimization**: miners compete to find the best configurations for four industry-standard variant calling tools (GATK, DeepVariant, bcftools, and freebayes). Every tool has hundreds of tunable settings

that affect accuracy, and the optimal configuration varies by genomic context, coverage depth, and variant type. Most labs never systematically explore this space, Minos does it continuously, approximately 20 rounds per day, with real economic incentives driving the search.

But hyperparameter tuning is only the starting point. Once the network has exhausted the configuration space of existing tools, Minos transitions to a **custom code submission model**: miners write and submit their own variant calling algorithms, novel neural networks, hybrid statistical-ML pipelines, region-specialized callers, in any language. The same blind evaluation infrastructure that benchmarks configurations will benchmark entirely novel approaches, creating an open arena where anyone in the world can build a better variant caller and prove it works against real genomic data.

The sections below describe how the current system operates: the three roles involved, how challenges are generated, how miners participate, and how results are scored.

2.1 Three Roles

Minos has three distinct roles. Figure 4 shows how they interact in a continuous loop.

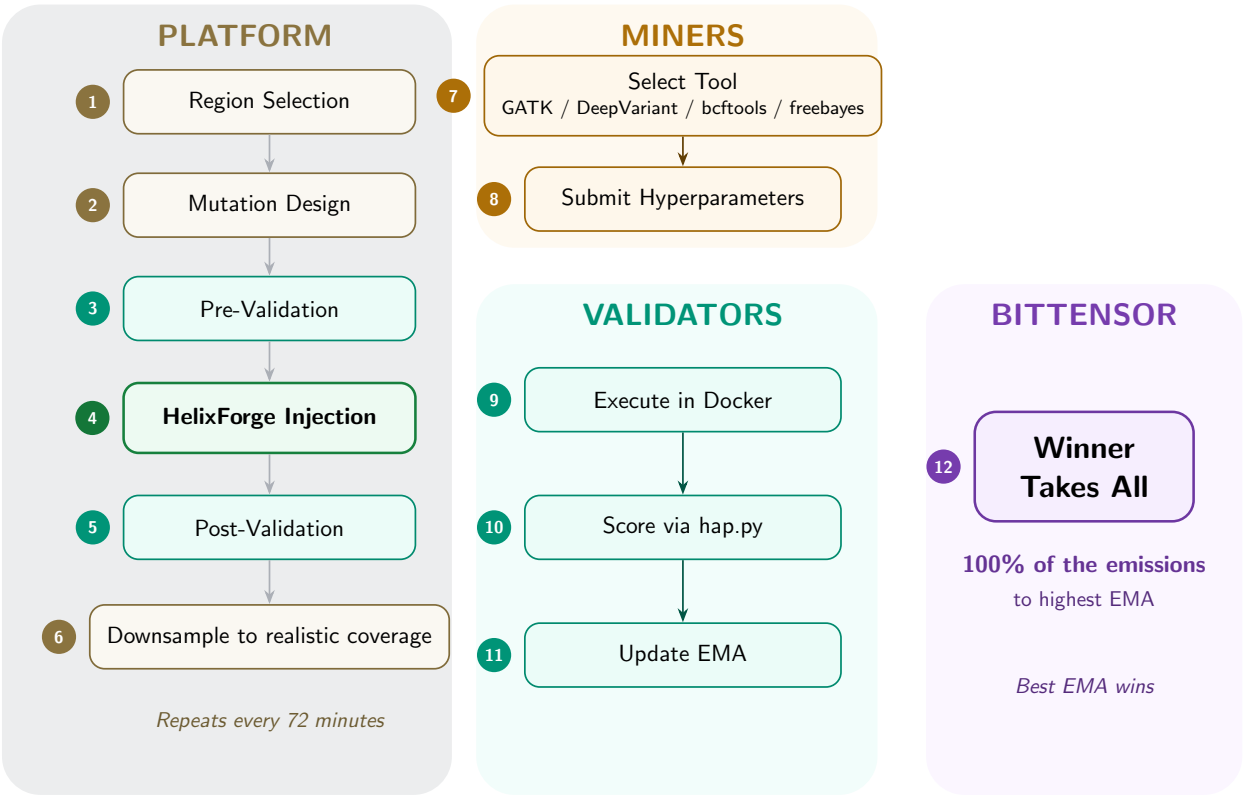

Figure 4: Detailed Minos architecture. **Left**: The platform generates fresh challenges every 72 minutes by injecting synthetic mutations via HelixForge and downsampling to random coverage. **Top-right**: Miners select a tool and submit optimized hyperparameters. **Center-right**: Validators execute in isolated Docker containers, score with hap.py, and update EMA. **Far right**: The highest EMA miner receives 100% of the emissions. Numbered badges show the order of operations; the cycle repeats every 72 minutes.

- **Platform (subnet owners)**: Every 72 minutes, the platform creates a new challenge by taking real human sequencing data and injecting hidden synthetic DNA mutations using **HelixForge**,

a mutation injection pipeline developed by the Minos team. These mutations are embedded directly into the raw sequencing reads so they are indistinguishable from real biological mutations.

- **Miners**: Miners choose one of four supported variant calling tools (GATK, DeepVariant, bcftools, or freebayes) and submit an optimized set of hyperparameters: settings that control how the tool processes the data. Miners do not run the tool themselves; they submit configurations for validators to execute.
- **Validators**: Validators receive the challenge genome and the miner's configuration, execute the variant calling tool inside an isolated Docker container, and score the output by comparing it to the known ground truth. Scores are submitted to the Bittensor blockchain.

2.2 Challenge Quality Pipeline

Not every mutation makes it into the final challenge. The platform uses a rigorous multi-stage quality pipeline to ensure that only confirmed mutations enter the truth set. Figure 5 shows how candidates are progressively filtered.

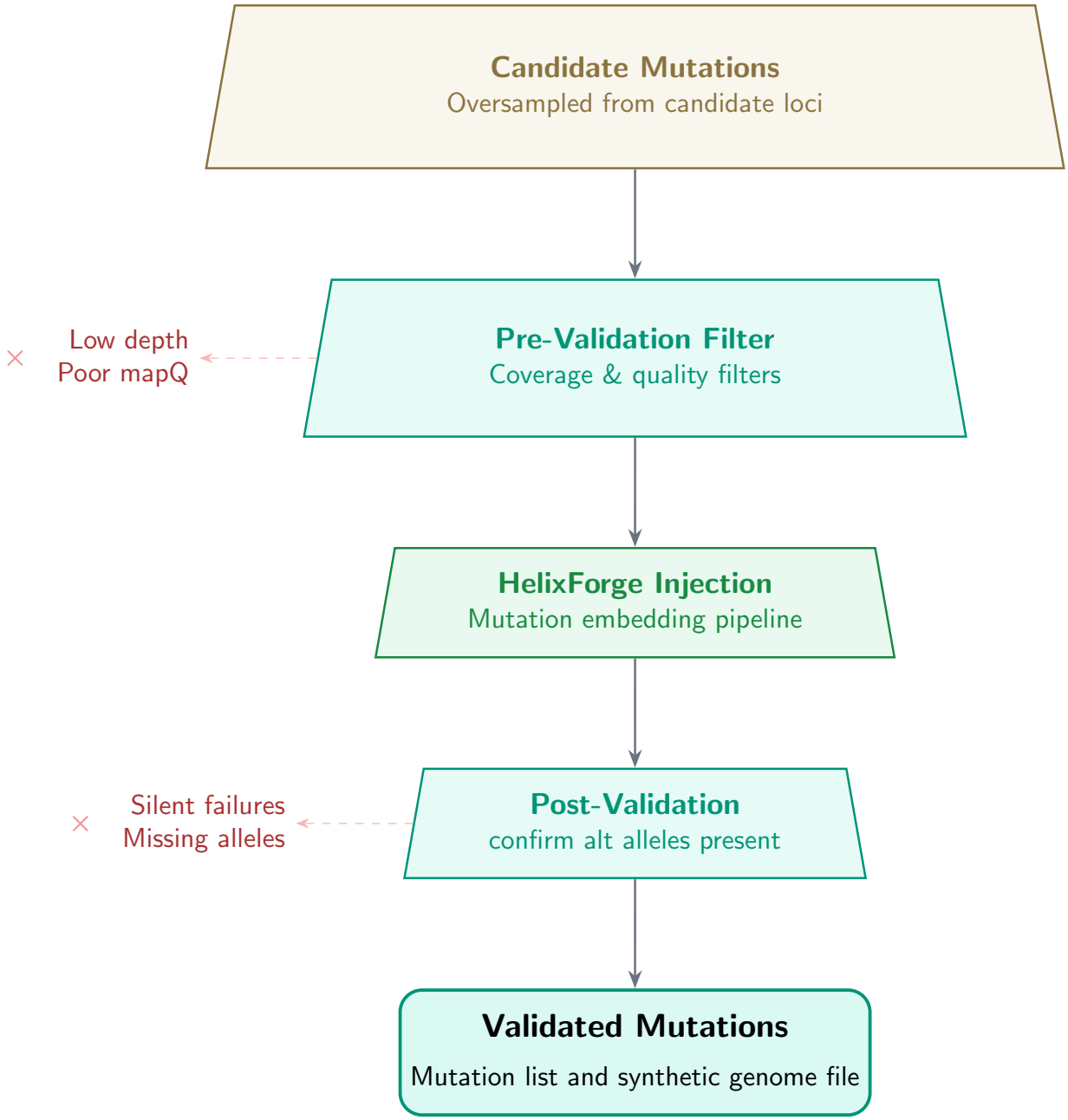

Figure 5: The validation funnel. Candidates are oversampled, then filtered through pre-validation (coverage and quality thresholds), HelixForge injection, and post-validation (alt allele confirmation). Only doubly-validated mutations enter the truth set.

2.3 Key Design Properties

- **Fresh challenges every round**: Every 72 minutes, the platform generates new synthetic mutations at random positions, so miners cannot memorize or cache answers.
- **Validator-side execution**: Miners submit configurations, not results. Validators execute the tool, eliminating self-reporting bias.

- **Multiple submissions allowed**: Miners can refine their configuration during the round window. Only the final submission is scored.
- **Deterministic**: The same configuration run by any validator on the same input produces identical results.

3 Roadmap

Minos is being built in five phases, and the strategic logic is deliberate. In Phase 1, miners compete to squeeze every last drop of accuracy out of existing, well-understood tools, GATK, DeepVariant, bcftools, and freebayes. These are the same tools used by Google, the Broad Institute, and thousands of clinical labs worldwide. By the time the network exhausts the optimization space of these established tools, Minos will have a battle-tested infrastructure: validated challenge generation, reliable scoring, proven incentive mechanics, and a growing database of benchmarked genomes.

That foundation is what makes Phase 2 transformative. Once existing tools are squeezed to their limits, Minos opens the floodgates: miners submit their *own* variant calling algorithms, novel neural networks, hybrid statistical-ML methods, region-specialized callers, evaluated on the same blind benchmarks. This is where Minos will truly shine. Instead of four tools competing on hyperparameters, the network becomes an open arena where anyone in the world can build a better variant caller and prove it works, every 72 minutes, against real data. The remaining phases scale this from small variants to structural variants, build a consensus AI from the best submissions, and deploy it as a production genomics service. Figure 6 shows the full progression.

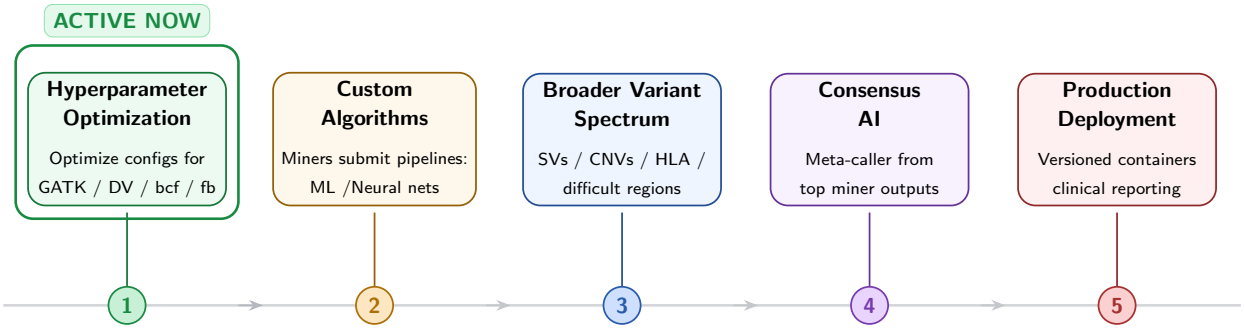

Figure 6: Minos development roadmap. Phase 1 (active, highlighted) validates the core infrastructure with hyperparameter optimization. Subsequent phases expand from configuration submission to custom code, broader variant types, consensus AI, and production deployment.

Phase 1: Hyperparameter Optimization (Active). Miners optimize configurations for four established variant calling tools: GATK HaplotypeCaller, DeepVariant, bcftools, and freebayes. This phase validates the full infrastructure: challenge generation, HelixForge injection, hap.py scoring, EMA tracking, and Winner-Take-All weight distribution. The output is a continuously updated leaderboard of optimal hyperparameters for each tool.

Phase 2: Custom Genotyping Algorithms. Miners submit their own variant calling code and pipeline instead of just configurations. This unlocks algorithmic innovation: neural network callers, hybrid methods, and region-specialized algorithms. Code executes in sandboxed Docker containers. An **innovation boost** mechanism will temporarily amplify rewards for miners who achieve breakthrough accuracy gains that exceed the network-wide best.

Phase 3: Broader Variant Spectrum. Minos expands beyond small variants (SNPs and indels) to include structural variants (large deletions, duplications, inversions), copy number variants, and clinically important regions like HLA genes. Scoring evolves to handle these more complex variant types.

Phase 4: Consensus AI. Aggregate the best-performing miner outputs into a unified consensus model that combines the strengths of diverse approaches. This "meta-caller" is expected to exceed the accuracy of any individual miner's pipeline.

Phase 5: Production Deployment. Ship the consensus caller as a production-grade service: versioned Docker images, automatic updates from subnet learnings, clinical-friendly reporting, and state-of-the-art accuracy tracking.

4 Incentive Mechanism

This section details the incentive mechanism developed and deployed by Minos. We descirbe how miner accuracy is measured, how scores are smoothed over time, and how alpha emissions are distributed.

4.1 Accuracy-Based Rewards

The core of the incentive mechanism relies on the accuracy of miner's performance. Miners are reward are assessed based on the F1 score of their config, which balances precision and recall in genomic data analysis.

Precision and Recall. After a validator runs a miner's configuration, the output is compared to the ground truth using **hap.py**, the industry-standard genomic benchmarking tool. For each variant type $v \in \{\text{SNP}, \text{INDEL}\}$:

$$\text{Precision}_v = \frac{TP_v}{TP_v + FP_v}, \qquad \text{Recall}_v = \frac{TP_v}{TP_v + FN_v}$$

where TP_v, FP_v, and FN_v are true positives, false positives, and false negatives respectively, as computed by hap.py against the merged truth set.

F1 Score. The harmonic mean of precision and recall:

$$F1_v = 2 \cdot \frac{\text{Precision}_v \cdot \text{Recall}_v}{\text{Precision}_v + \text{Recall}_v}$$

In plain terms:

- **Precision**: Of all the mutations the miner's tool reported, what fraction were real? (Fewer false alarms = higher precision.)
- **Recall**: Of all the real mutations in the data, what fraction did the miner's tool find? (Fewer missed mutations = higher recall.)
- **F1 Score**: The balanced combination of precision and recall. A perfect F1 of 1.0 means every real mutation was found and nothing was falsely flagged.

4.2 Multi-Component Scoring

Rather than using a simple weighted F1, Minos employs a **four-component scoring system** that ensures balanced performance across all dimensions of variant calling quality:

$$S = w_1 \cdot C_{\text{core}} + w_2 \cdot C_{\text{complete}} + w_3 \cdot C_{\text{FP}} + w_4 \cdot C_{\text{quality}}, \quad w_1 \gg w_2,\, w_3 > w_4, \quad \sum w_i = 1$$

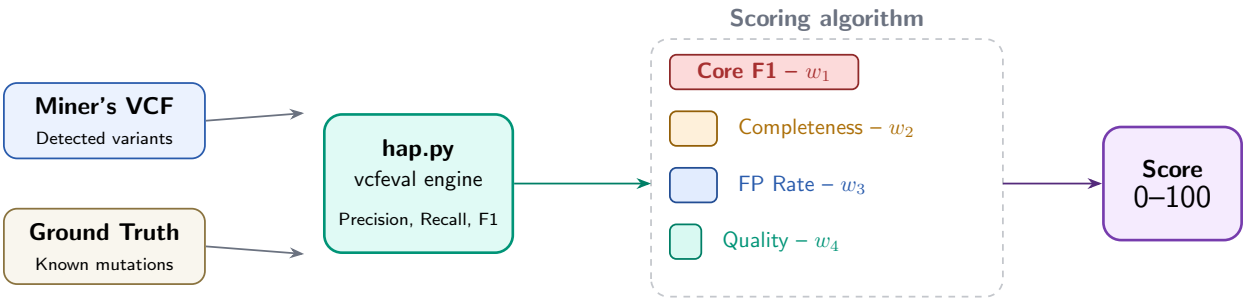

Figure 7: Scoring pipeline. The miner's variant calls are compared to ground truth via hap.py. The scoring algorithm weights core F1 accuracy most heavily, with completeness, false positive rate, and biological quality ratios providing additional differentiation. Exact weights are tuned to reward balanced variant calling performance.

Each component is computed from hap.py metrics as follows:

Emphasis Function. Several components use a nonlinear emphasis function that severely penalizes incomplete performance:

$$E(x, \gamma) = 1 - (1 - x)^{\gamma}$$

Higher γ values make the penalty more severe near zero.

Core Component (w_1). Truth-weighted F1 score with nonlinear emphasis:

$$C_{\text{core}} = E\left(\frac{F1_{\text{SNP}} \cdot N_{\text{SNP}} + F1_{\text{INDEL}} \cdot N_{\text{INDEL}}}{N_{\text{SNP}} + N_{\text{INDEL}}},\ \gamma = 0.5\right)$$

where N_{SNP} and N_{INDEL} are the truth set variant counts. When truth set counts are unavailable, the fallback weighting is $0.7 \cdot F1_{\text{SNP}} + 0.3 \cdot F1_{\text{INDEL}}$.

Completeness Component (w_2). Average recall and genomic coverage:

$$C_{\text{complete}} = \frac{1}{2}\left[E\left(\frac{R_{\text{SNP}} + R_{\text{INDEL}}}{2},\ \gamma = 3.0\right) + E(1 - f_{\text{NA}},\ \gamma = 2.0)\right]$$

where R_v is recall for variant type v, and $f_{\text{NA}} = \max(f_{\text{NA,SNP}},\ f_{\text{NA,INDEL}})$ is the worst-case fraction of uncallable regions across variant types.

False Positive Component (w_3). Exponential penalties for excessive false positives:

$$C_{\text{FP}} = \frac{1}{2}\left[\exp\left(-\frac{\max(0,\ r_{\text{FP}} - t_{\text{FP}})}{t_{\text{FP}}}\right) + \exp\left(-\frac{|r_{\text{size}} - 1|}{0.10}\right)\right]$$

where $r_{\text{FP}} = \frac{FP_{\text{SNP}}+FP_{\text{INDEL}}}{Q_{\text{SNP}}+Q_{\text{INDEL}}}$ is the false positive rate, $r_{\text{size}} = \frac{Q_{\text{SNP}}+Q_{\text{INDEL}}}{N_{\text{SNP}}+N_{\text{INDEL}}}$ penalizes calling significantly more or fewer variants than expected, and $t_{\text{FP}} = \max(0.002,\ 1/N_{\text{total}})$ is a dynamic target that scales with the evaluation set size.

Quality Component (w_4). Biological plausibility via Ti/Tv and Het/Hom ratios:

$$C_{\text{quality}} = \frac{1}{2}\left[\overline{P}_{\text{TiTv}} + \overline{P}_{\text{HetHom}}\right]$$

where each penalty term is computed as:

$$P(\delta, \sigma) = \exp\left(-\frac{|\delta|}{\sigma}\right)$$

with $\sigma = 0.1$ for Ti/Tv ratios and $\sigma = 0.15$ for Het/Hom ratios, where δ is the difference between query and truth values. If a ratio is unavailable (e.g., zero truth variants), the penalty defaults to 1.0 (no penalty).

The final score is scaled to a 0–100 range: $S_{\text{final}} = 100 \cdot S$.

Gaming Resistance. This design prevents miners from optimizing a single metric:

- **Truth-weighted F1**: Adapts to actual variant distributions rather than fixed weights
- **Emphasis functions**: Nonlinear penalties make zero performance in any area catastrophic
- **Independent components**: Strong SNP performance cannot compensate for failed INDEL calling
- **Multiple orthogonal metrics**: Recall, FP rate, and quality ratios must all be reasonable

4.3 Score Smoothing (EMA)

A single round can have random variance, so Minos uses an **Exponential Moving Average (EMA)** to smooth scores over time:

$$\text{EMA}_t = (1 - \lambda) \cdot \text{EMA}_{t-1} + \lambda \cdot S_t$$

where $\lambda = 0.1$ gives 10% weight to new scores and 90% to historical performance.

- All miners start with EMA = 0. On the first scored round, the update yields $\text{EMA}_1 = \lambda \cdot S_1$ (i.e., 10% of the first score). This means new miners must sustain high performance over multiple rounds to build a competitive EMA, preventing a single strong round from immediately dominating the leaderboard.
- The EMA converges to the true expected score with characteristic timescale $\tau = -1/\ln(1-\lambda) \approx 9.5$ rounds.
- With 72-minute task intervals, this corresponds to approximately 11.4 hours for full convergence.

- If a miner goes offline, their EMA **decays by 5% per missed round**, preventing stale scores from persisting on the leaderboard.

Figure 8 illustrates how EMA smooths individual round scores and how decay affects absent miners.

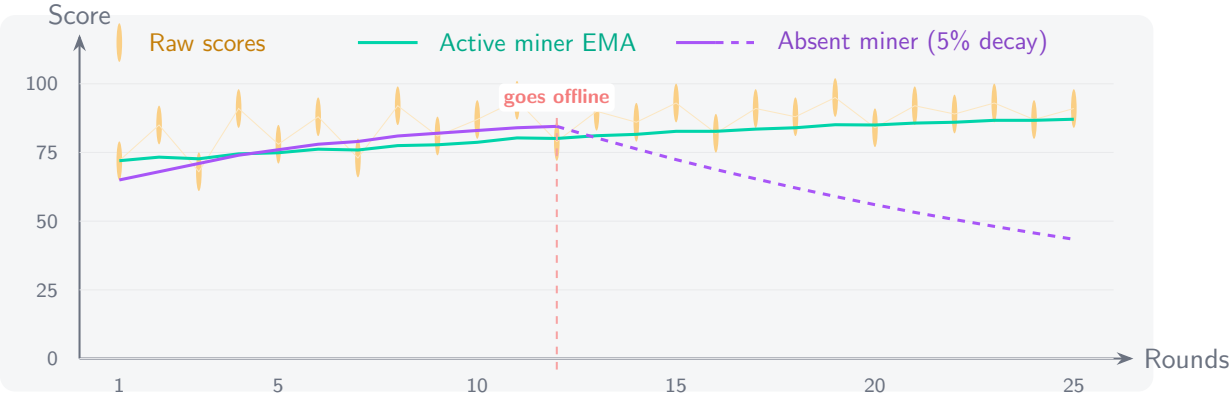

Figure 8: EMA smoothing and decay. The active miner (teal) builds a stable reputation over 25 rounds. A second miner (purple) goes offline at round 12, and their EMA decays by 5% per round, falling from 84.5 to 43.4 in 13 rounds.

4.4 Winner-Take-All Weight Distribution

The current implementation uses a **Winner-Take-All** model: the miner with the highest EMA score receives 100% of the emissions. Everyone else receives nothing. This creates maximum competitive pressure.

Winner Determination.

1. **Eligibility via Participation**: A miner must have been scored in at least $R_{\min} = 10$ rounds to be eligible for weights.

2. **Score comparison**: EMA scores are compared with a tolerance of 10^{-9} for epsilon-based tie handling.

3. **Tie-breaking**: If multiple miners share the highest EMA score within this tolerance, the miner with the **earliest submission timestamp in the most recent round** wins. This rewards responsiveness and consistent engagement rather than registration seniority.

Two-Phase Distribution. Let $\mathcal{E} = \{m \in \mathcal{M} : \text{rounds}(m) \geq R_{\min}\}$ be the set of eligible miners. Weight distribution operates in two phases:

Warm-up phase ($\mathcal{E} = \varnothing$): Before any miner reaches $R_{\min}$ scored rounds, emissions are split among the **top-3 active miners** by EMA score:

$$w_m = \begin{cases} 0.50 & \text{1st place by EMA} \\ 0.30 & \text{2nd place by EMA} \\ 0.20 & \text{3rd place by EMA} \\ 0.0 & \text{otherwise} \end{cases} \quad \text{(warm-up)}$$

This rewards early participants proportionally to their performance while still distributing risk across multiple miners during the bootstrapping period.

Competitive phase ($\mathcal{E} \neq \varnothing$): Once at least one miner reaches eligibility, the system transitions to Winner-Take-All:

$$w_m = \begin{cases} 1.0 & \text{if } m = \arg\max_{m' \in \mathcal{E}} \text{EMA}_{m'} \\ 0.0 & \text{otherwise} \end{cases}$$

Emissions are distributed proportionally to these weights:

$$\text{Reward}_m = R_{\text{total}} \cdot w_m$$

Rationale. The warm-up phase ensures early miners are rewarded for joining and contributing scoring data, with the top-3 split maintaining competitive pressure even before full eligibility is reached. Once sufficient history exists, Winner-Take-All activates, concentrating all rewards on the single best performer and maximizing competitive pressure: miners must be the absolute best, not merely good, to earn emissions.

5 Why Miners are Incentivized to Not Cheat

The core security mechanism is elegantly simple: **miners don't know which mutations in the challenge are synthetic**.

5.1 Synthetic Mutation Injection

Every round, the platform uses **HelixForge**, a mutation injection pipeline developed by the Minos team, to embed synthetic DNA mutations directly into real sequencing reads. These mutations:

- Are embedded at the **read level** (not just annotated), so they are indistinguishable from real biological mutations at the algorithmic level.
- Follow **realistic biological patterns** that make them indistinguishable from real mutations at the algorithmic level.
- Are placed at **random positions** that change every round, so previous answers are useless.
- Are **pre-validated** (confirmed the injection site has sufficient data quality) and **post-validated** (confirmed the mutation actually appears in the output).

5.2 The Blind Evaluation Framework

A critical design property is that evaluation is **restricted to the genomic loci containing synthetic mutations**, while miners operate with **no knowledge of which positions are synthetic**. This asymmetry creates a blind evaluation framework analogous to masked assessment protocols in clinical trials.

Figure 9 illustrates this information asymmetry.

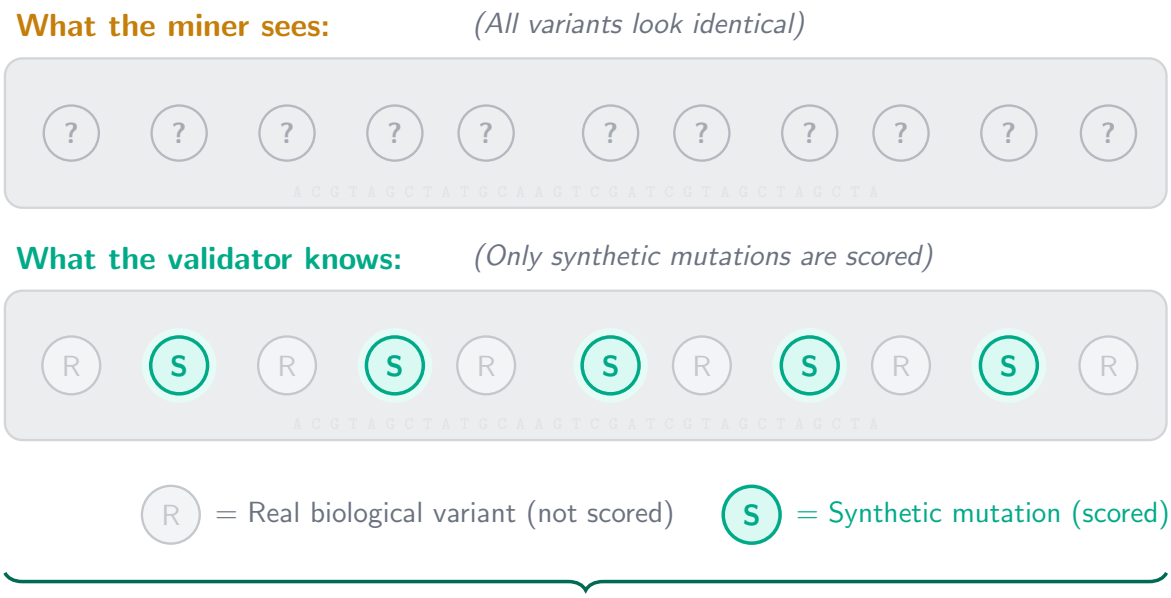

Figure 9: Blind evaluation framework. **Top**: Miners see a genome region where all variants are indistinguishable. **Bottom**: Validators know which are synthetic and score only those positions. The only winning strategy is genuine, high-quality variant calling.

Why this works:

- **Scoring only covers synthetic mutations**: Validators evaluate only the synthetic loci, not the entire region. This gives a clean, controlled signal.
- **Miners can't tell which are synthetic**: The mutations are embedded in real reads and are biologically realistic. There is no signal to distinguish them.
- **The only winning strategy is genuine accuracy**: A miner that does high-quality variant calling across the entire region will naturally detect the hidden synthetic mutations. Cutting corners means missing them.

5.3 Mutation Space Analysis

Even for a small challenge, the number of possible mutation configurations is astronomically large, rendering any memorization or pre-computation attack mathematically infeasible.

Total Mutation Space. For M_{SNP} SNPs and M_{indel} indels injected into a genomic window with P candidate positions (where $P = W \cdot \phi$ for window size W and confident region fraction ϕ), the total number of possible mutation configurations is:

$$\Omega = \binom{P}{M_{\text{SNP}} + M_{\text{indel}}} \cdot A_{\text{SNP}}^{M_{\text{SNP}}} \cdot A_{\text{indel}}^{M_{\text{indel}}}$$

where A_{SNP} is the number of possible alternate bases per SNP position and A_{indel} reflects insertion/deletion allele diversity.

Taking the logarithm:

$$\log_{10}(\Omega) = \log_{10}\binom{P}{M} + M_{\text{SNP}} \cdot \log_{10}(A_{\text{SNP}}) + M_{\text{indel}} \cdot \log_{10}(A_{\text{indel}})$$

Entropy Bound. The information-theoretic entropy of the mutation space in bits:

$$H = \log_2(\Omega) = \frac{\log_{10}(\Omega)}{\log_{10}(2)} \approx 3.32 \cdot \log_{10}(\Omega)$$

This entropy represents the minimum number of bits required to specify a particular mutation configuration, and equivalently, the amount of information a miner would need to pre-compute all possible answers.

Key Results. For representative Minos challenge configurations, the mutation space exceeds 10^{60} possible combinations—dwarfing the number of atoms in the observable universe ($\sim 10^{80}$). Even the smallest practical configuration exceeds 10^{30} possible mutation sets, which surpasses:

- The number of atoms in the human body ($\sim 10^{28}$)
- Total compute operations in human history ($\sim 10^{24}$)
- Any conceivable lookup table storage capacity

A typical Minos round produces a number of possible mutation sets so astronomically large that it renders any memorization or pre-computation attack mathematically infeasible.

Another advantage of synthetic mutation injection is that it enables evaluation across a mutation space far exceeding all current human genetic diversity. There are approximately 8 billion humans alive today, and roughly 117 billion humans have ever lived. Each human genome contains ∼4–5 million variants relative to the reference genome. HelixForge can generate synthetic genome configurations that vastly exceed the total genetic variation across all humans who have ever lived.

By continuously sampling from this effectively infinite mutation space, Minos tests variant calling pipelines against genomic variation that spans far beyond what has been observed in any human population. Miners must build algorithms that generalize to *all biologically plausible genomes*, not just the specific mutation patterns present in existing datasets.

The only viable strategy is to run a genuine, high-quality variant calling pipeline, exactly as intended.

6 What Minos Produces

Beyond the competitive benchmarking itself, Minos generates three valuable outputs that compound over time. Figure 10 illustrates how the competitive dynamics create a self-reinforcing cycle.

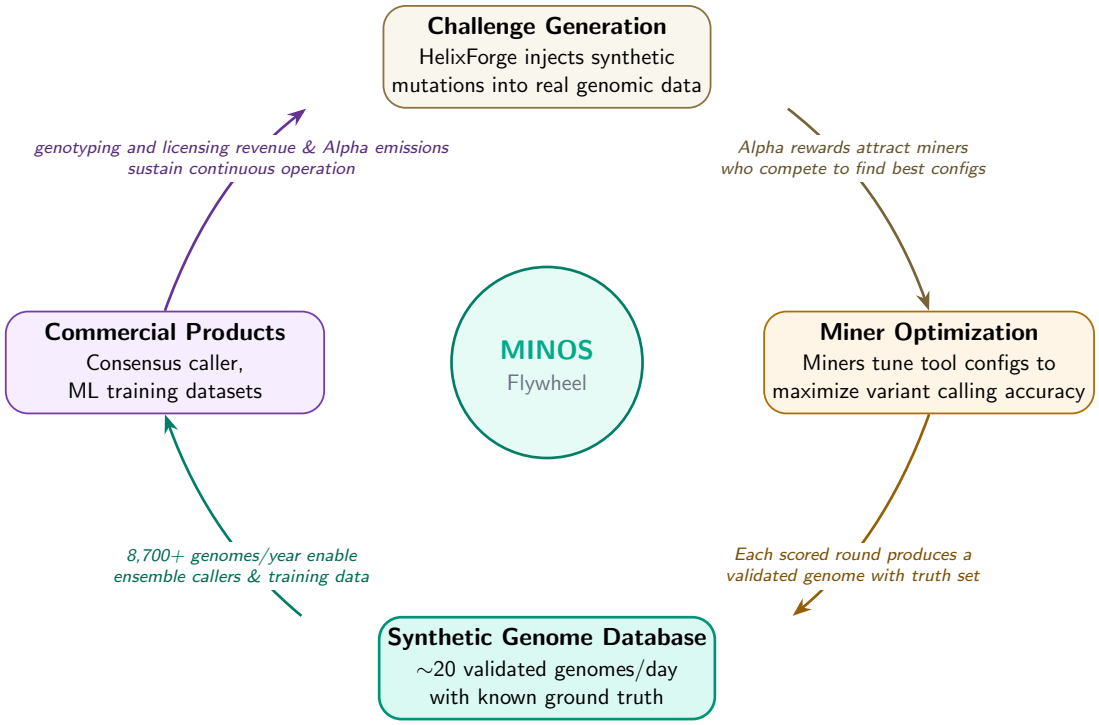

Figure 10: The Minos flywheel. Each 72-minute challenge produces a validated synthetic genome as a byproduct of scoring. These accumulate into a database that enables commercial products (consensus caller, AI-powered variant calling infrastructure, ML training data). Revenue sustains the platform, and the cycle accelerates with each round.

1. **World's Largest Validated Synthetic Genome Database.** Every round generates a new synthetic genome with known ground truth. At ∼20 genomes per day, this reaches over 7,300 validated genomes per year, each with confirmed ground-truth mutations. This database is a unique resource for training next-generation variant calling models.

2. **Consensus Variant Caller.** By aggregating the best-performing configurations across tools and genomic contexts, Minos can produce an ensemble caller that combines the strengths of multiple approaches.

3. **AI-powered variant calling infrastructure.** Labs, pharmaceutical companies, and biobanks can submit DNA sequencing samples and receive benchmarked, high-accuracy variant calls with full provenance.

6.1 Projected Database Growth

Figure 11 shows the expected growth of the synthetic genome database over the first three years.

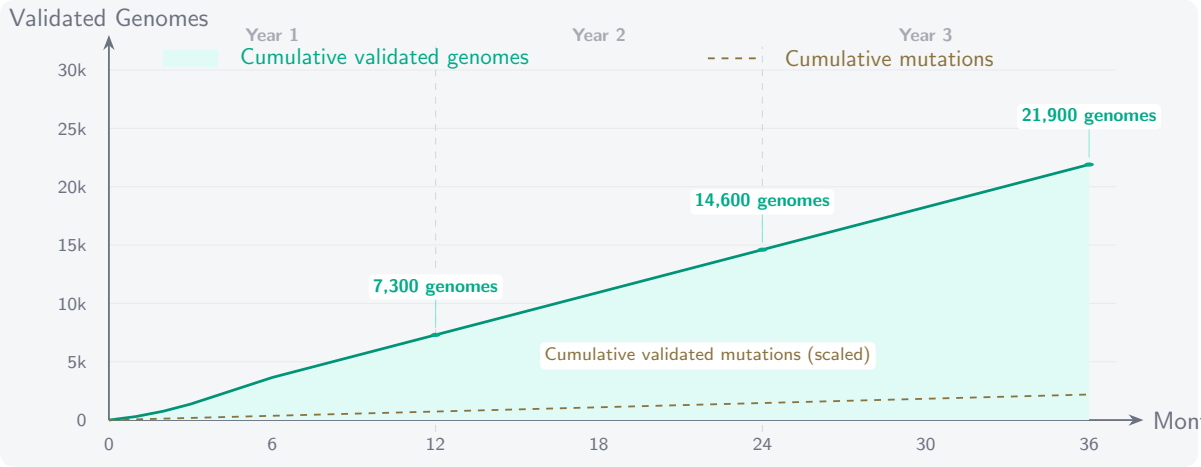

Figure 11: Projected database growth over 36 months. At ∼20 challenges per day, the database reaches ∼7,300 validated genomes in Year 1, ∼14,600 by Year 2, and ∼21,900 by Year 3, each containing validated synthetic mutations.

6.2 Expected Performance Improvement

Competitive optimization is expected to significantly outperform default tool configurations. Figure 12 illustrates the projected accuracy gains.

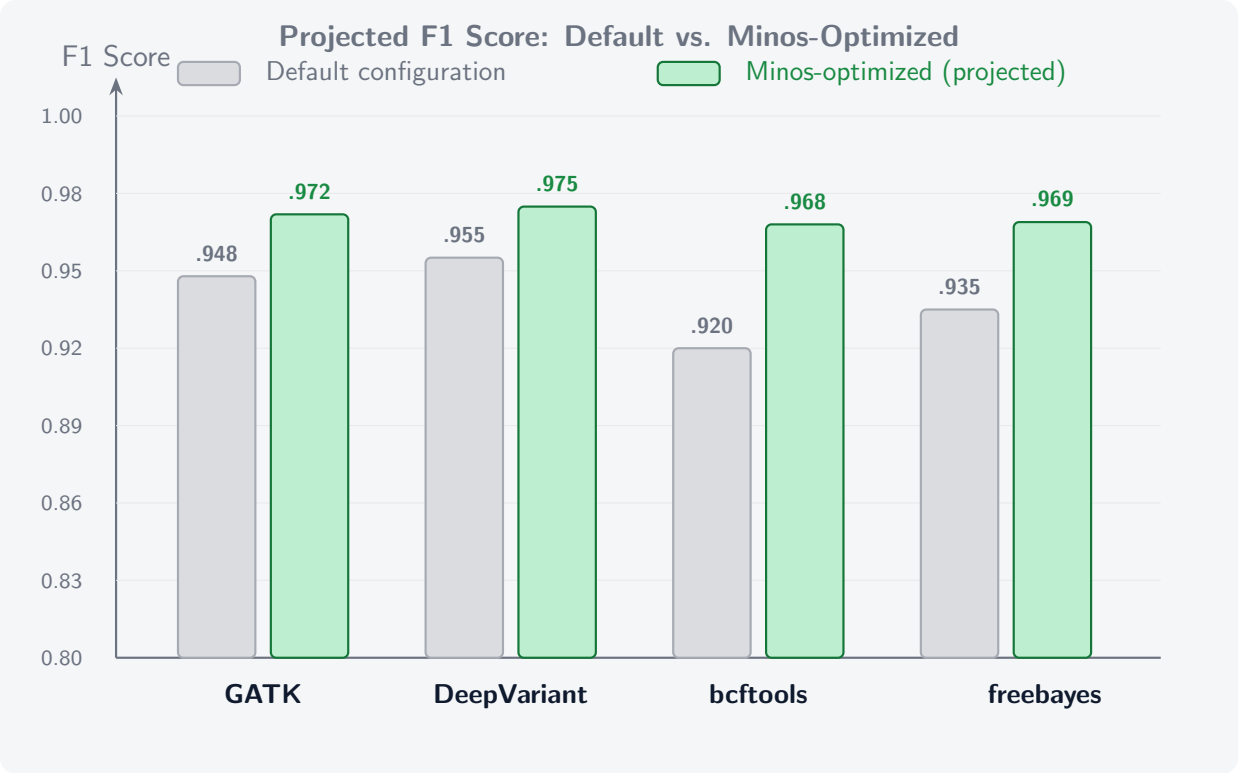

Figure 12: Projected F1 score improvement from continuous hyperparameter optimization. Default configurations (gray) represent published baseline performance. Minos-optimized values (green) reflect expected gains from competitive tuning. Even small F1 gains (2–5%) translate to thousands of additional correctly called variants per genome.

7 Conclusion

Minos brings a real-world computational biology problem to the Bittensor ecosystem: identifying DNA mutations with the highest possible accuracy. By combining established genomic tools with decentralized, incentive-driven optimization, the subnet creates a system where accuracy improvements are immediately rewarded, independently verified, and made available to the entire genomics community.

The opportunity is significant. Genomic sequencing is scaling from millions to billions of samples over the next decade, and every one of those samples needs accurate variant calling. Minos is building the infrastructure to make that happen: continuously benchmarked, transparently scored, and economically incentivized.

For miners, the path is clear: pick a variant calling tool, optimize your configuration, and submit it. The best performer gets all the emissions. As the subnet matures through its five phases, the scope expands from hyperparameter tuning to custom algorithms to a production genomics service, creating growing opportunities for innovation and reward.

References

[1] National Human Genome Research Institute. (2024). "The Cost of Sequencing a Human Genome." Available at: https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data

[2] Zook, J.M., et al. (2014). "Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls." *Nature Biotechnology*, 32, 246–251.

[3] Zook, J.M., et al. (2019). "An open resource for accurately benchmarking small variant and reference calls." *Nature Biotechnology*, 37, 561–566.

[4] Krusche, P., et al. (2019). "Best practices for benchmarking germline small-variant calls in human genomes." *Nature Biotechnology*, 37, 555–560.

[5] Ewing, A.D., et al. (2015). "Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection." *Nature Methods*, 12, 623–630.

[6] Population Reference Bureau. (2022). "How Many People Have Ever Lived on Earth?" Available at: https://www.prb.org/articles/how-many-people-have-ever-lived-on-earth/

[7] Downey, B., et al. (2023). "Bittensor: A Decentralized Neural Network." Bittensor Foundation Whitepaper. Available at: https://bittensor.com/