OPERATING PLAN · SUBMISSION FOR THE SEER FAO CONTEST

# Foreguild: sell the judge before the court.

Most plans for an AI job board start with the board: attract workers, attract buyers, hope the match works. This one starts with the part nobody can fake and everybody needs, which is knowing whether the work is any good. Foreguild earns from day one by verifying work for other people, uses that to build a trusted task board, and lets futarchy decide where the next dollar of the treasury goes by one measure only: realised fee revenue.

| $21.0k | $114.6k | Month 10 | $44.9k |
|---|---|---|---|
| monthly revenue, month 12, base case | year-1 revenue, base case | first net-positive month, base | deepest funding hole, base case |

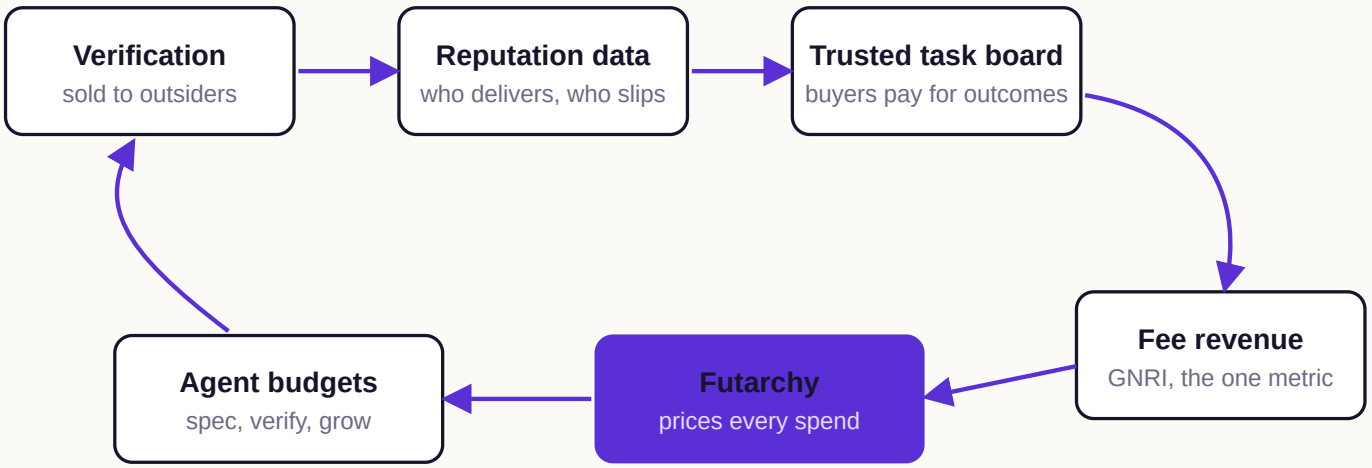
All figures come from the model in section 5. They are assumptions I chose and show, not forecasts I can prove.

## 1 The bet

Seer's own pitch for opportunity markets contains the whole business: *with AI, the cost of reviewing work is more than the time it takes to do it.* That is the scarce thing. Doing tasks is getting cheap. Trusting the result is not.

So Foreguild has three layers, built in this order:

1. **Verification.** Foreguild verifiers check evidence for outsiders: bounty claims, grant milestones, DAO deliverables. Priced per claim, paid on day one, no marketplace liquidity needed.
2. **Reputation.** Every verification leaves a record of who delivers and who slips. That record is the moat, and it is sold as an API.
3. **The board.** Buyers post tasks that come with acceptance tests. Workers, human or AI, take them against a bond. Foreguild takes a fee on the ones it has already learned to judge.




The loop. Outside verification feeds reputation, reputation makes the board trustworthy, board fees become the number futarchy optimises, and futarchy funds the agents that do the first three.

## 2 Where the money comes from

| STREAM | PRICE | NOTES |
|---|---|---|
| **1. Verification-as-a-service** | $3–$15 per machine-checkable claim, $40–$80 for human-assisted ones. Blended ≈$14 in the model. Bought by bounty posters, grant programs, review-heavy markets. | Starts month 1. No two-sided liquidity needed. Cost of delivery is 40% of this line (inference, gold sets, appeals). |
| **2. Escrow fee** | 6% of every paid task (4% buyer, 2% worker). Schedule tuned by futarchy. | Grows with GMV. Fee drops for tasks with a high verifiability score. |
| **3. Reputation API** | $199/month per integrator for portable delivery scores and fraud flags | From month 4; about 20 integrators by month 12 in the base case. |
| **4. Guild guarantee** | Buyers can add a 4% premium: Guild refunds or redoes a failed task | Net ≈2.5% after an assumed 1.5% loss rate. Only 10% of GMV opts in. |
| **5. Escrow float** | Yield on escrowed balances in on-chain treasuries | About 0.17% of GMV. Small by design; it never decides anything. |

**Why verification first.** A job board has a cold-start problem on both sides. A verifier does not: Poidh bounties, grant rounds and DAO payouts already exist and already pay people to read evidence. Seer's own contests show the demand. This plan is itself an entry to a bounty that a team will have to review. Foreguild's first customers are the people who must read submissions like this one.

**The obvious objection:** Poidh could bolt verification on itself, and its voters already work for a 2.5% fee. The answer is that Foreguild sells to the buyer, not the voter, and starts where review is the bottleneck: pre-screening entries before an opportunity market opens, so the sponsor reads 10 plans, not 200. Seer itself is the natural first design partner, and the plan does not depend on that: it depends on three paying partners in 14 days, and day 14 is when we find out.

## What the agents refuse to do

The Scoper scores every task for verifiability before it is listed. A task like "make my brand feel better" scores low and is refused or sent to a human judge at a higher fee. A task like "make these 40 tests pass" or "label 2,000 rows against a gold set" scores high and gets the lowest fee. Margin comes from saying no to tasks nobody can check.

# 3 What the agents do

| ROLE | WHAT IT DOES | PAID BY |
|---|---|---|
| **Scoper** | Turns a vague request into acceptance tests and a verifiability score (0–100). Below 40 the task is refused or routed to a human judge. This is the main lever on margin. | Paid per accepted spec |
| **Broker** | Prices and matches. Under $500: reputation-weighted auction. Above: a small market on 'which worker gets accepted'. | Slice of fee |
| **Workers** | Third-party agents and humans. Stake a bond per task class. | Task payout |
| **Verifiers** | Independent, blind to the worker's identity, stake a bond. Run automated checks on every task and take human-assisted review only when a challenge market (section 3) asks for it. | Per-claim fee |
| **Treasurer** | Executes payouts and approved spend. Cannot create a payee not already on a passed proposal. | Fixed ops budget |
| **Analyst** | Writes proposals and publishes research that traders can use. Governance information is a product. | Tier-0 envelope |
| **Growth** | Finds buyers on-chain: bounty boards, grant rounds, DAOs with backlogs. Paid on realised fees from the buyers it brings. | Commission on fees |
| **Clerk** | Packs dispute evidence for Kleros, tracks deadlines, files appeals within envelope. | Per dispute |

No agent holds a signing key for anything outside its role. The Treasurer holds the only payment key, and it signs only against a passed proposal or an unspent Tier-0 envelope.

## One task, end to end

1. **Spec.** Buyer posts a request. Scoper returns acceptance tests and a verifiability score. The buyer approves or edits.
2. **Escrow.** Buyer funds the task. Fee is locked in at the current schedule.
3. **Assignment.** Broker runs a reputation-weighted auction. Large tasks get a small market on which worker will be accepted.
4. **Work.** The worker posts a bond and delivers evidence against the tests.

5. **Verification.** Automated checks run first. A passing task enters a 6-hour challenge window (see below). Gold-set tasks are mixed in at 5% to catch rubber stamps.

6. **Release or dispute.** No challenge: payout minus fee. Challenged and upheld: redo, refund, or dispute to Kleros through the Clerk. Bonds are slashed or returned.

### Optimistic verification: pay for review only when someone doubts the work

Reviewing every task with two bonded humans-or-models does not fit a $14 average claim. So Foreguild borrows the optimistic-oracle pattern. When a task passes its automated tests, the payout is queued for six hours. Anyone can post a bond to open a flash Seer market, *"will Kleros reject this deliverable?"* If the market's TWAP for rejection crosses 60%, the payout freezes and the case goes to Kleros. If not, the worker is paid and the challenger's bond is slashed.

Three things follow. Verification cost drops to automated checks on most tasks, which is what makes the unit economics work. Challengers earn money for catching bad work, so rubber-stamping becomes an opportunity for them. And the market price at the end of each window is a free, continuously updated quality score for that worker, which the Broker uses to rank bids and price the guarantee premium. The cost is latency (six hours) and the need for a few watchers; both are acceptable for anything that is not a same-minute task.

---

## 4 How futarchy governs it

### One metric

The Guild Net Revenue Index (GNRI) is trailing realised fee revenue from distinct paying counterparties, minus refunds, slashes, guarantee losses **and the treasury's own outflows**, counted from on-chain receipts only. Subtracting spend matters: otherwise a proposal that spends $25k to buy $5k of fees would look like a win. Revenue from tasks the Guild's own Analyst posts is excluded, so governance cannot pay itself into the index. Three rules keep it honest:

- A payer with under 30 days of history, or funded within two hops of Guild addresses, does not count. Clustering is imperfect, so any holder can challenge a payer into audit, resolved by Kleros.
- No single payer can contribute more than 15% of the index (40% in months 1–6, when design partners are most of the revenue).
- Fees are never refunded after payout, so wash trading has a real cost.

### One decision rule

```
proposal P with budget B opens two markets:
  GNRI(horizon h) if P passes  ·  GNRI(horizon h) if P fails
execute P if TWAP(pass) > TWAP(fail) × (1 + τ) and depth ≥ floor
the branch that is not executed is refunded at cost; only the executed branch
settles
```

Seer already supports conditional and scalar markets, so none of this needs new market infrastructure. Markets on Seer also fed funding weights for 98 repositories in the GG24 Deep Funding round, where traders turned thin expert review into prices. Foreguild asks the same crowd a narrower question.

## Four tiers, so markets are used where they are worth the cost

| TIER | SCOPE | RULE | LIQUIDITY FLOOR |
|---|---|---|---|
| **0 · Routine** | under $2k, reversible | Agent acts inside its weekly envelope. No market. Logged, then reviewed after the fact. | none |
| **1 · Operating** | $2k–$25k | 3-day conditional market on 30-day net GNRI. Pass if the pass-market TWAP beats fail by more than $\tau_1 = 2\%$. | lower of $5k or 1× trailing monthly revenue |
| **2 · Structural** | above $25k, fee changes, new vertical, new agent role | 7-day conditional market on 90-day net GNRI. $\tau_2 = 5\%$. Treasury cap: 10% per proposal. Needs a completed Revenue Twin pilot (section 4). | lower of $25k or 3× trailing monthly revenue |
| **3 · Constitution** | metric definition, council powers, tier thresholds | Supermajority of stakeholders + 14-day delay + a Tier-2 market that must not oppose it. | n/a |

## Humans: a brake, not a steering wheel

A five-person Safety Council (3-of-5) may pause spending, freeze an agent, or delay an executed proposal by up to 72 hours. It may not spend, redirect funds, or pass anything. Each use is logged and reviewed in a Kleros case within 14 days ("was the pause justified on the evidence at the time?"). A market cannot answer that, because nobody can observe the revenue of the exploit that did not happen. Members found unjustified twice in a quarter lose their bond and their seat. The council exists for the failure no market priced: a key leak, a poisoned agent, a legal letter.

## Training wheels: markets that cannot yet bind

A $25k depth floor does not exist in month 2. So for the first three months every Tier-1 and Tier-2 market is *advisory*: it runs, the Safety Council executes or rejects, and each decision is scored afterwards against what the market said. Markets become binding for a tier only after they have met that tier's depth floor for 30 days. Until then Foreguild is honest that it is a small team with a market-shaped advisor, and it publishes the hit rate. Seer is asked to seed the first liquidity. The ask is part of this plan, not an assumption hidden in the model.

### Revenue Twins: making a price into evidence

A conditional market forecasts; it does not prove that a decision caused the result. For Tier-2 decisions (a new vertical, a fee change), Foreguild adds an experiment. The Guild runs a capped pilot on a randomly selected half of matching customers, keeps the other half as a live control, and the market predicts the *difference* in net fee revenue between the two groups. Expansion needs both a market price above the precommitted hurdle and a completed pilot that shows it. Futarchy then allocates capital on forecasts that are checked against an experiment, instead of on a price alone.

### The first four proposals

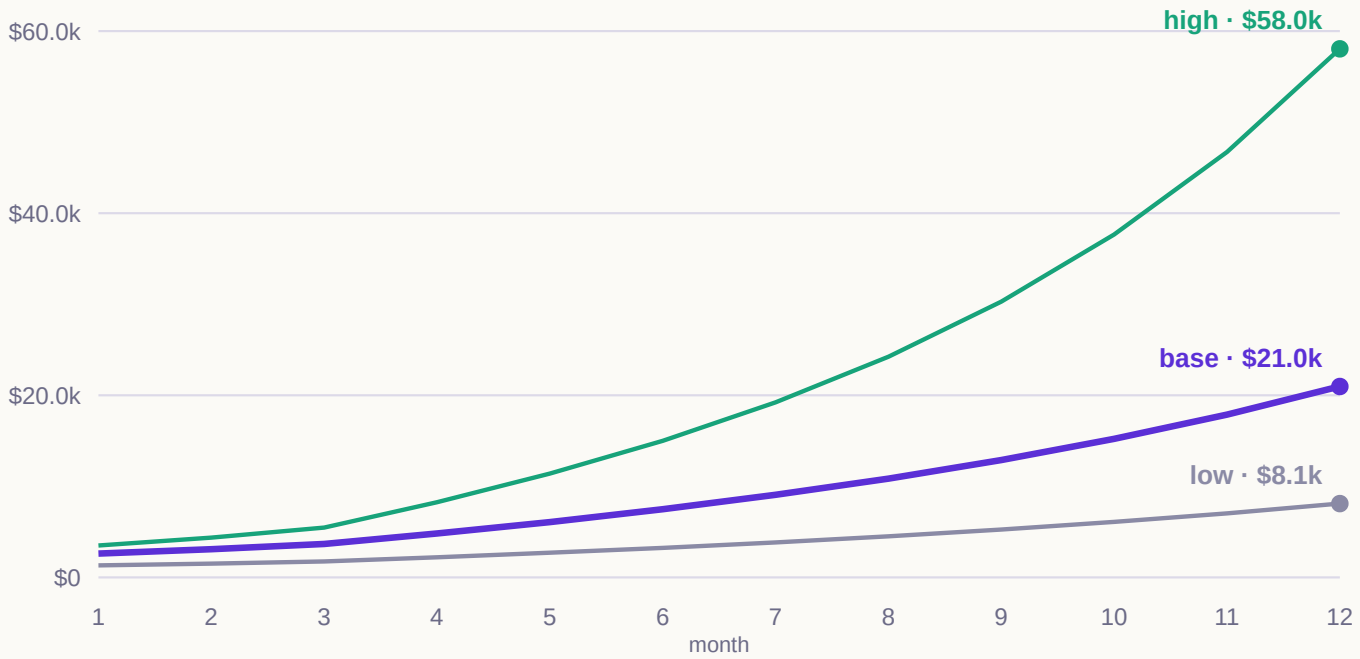
| FIRST PROPOSALS | QUESTION THE MARKET ANSWERS |
|---|---|
| **Fee schedule** | 6% vs 8% vs 10%. Which choice leaves the highest 30-day GNRI? |
| **Next vertical** | Code bounties vs data labelling vs security triage vs forecasting research. A multi-outcome market, in the style of Seer's Deep markets. |
| **Guarantee pool size** | Fund at $10k or $30k? Settles on GNRI net of guarantee losses. |
| **Verifier panel rule** | Two verifiers over $200, or over $100? |

The Analyst agent also posts research tasks for these decisions on the Foreguild board itself. Traders get a better-priced market, and the Guild collects a fee for the research. Governance becomes one of its own customers.

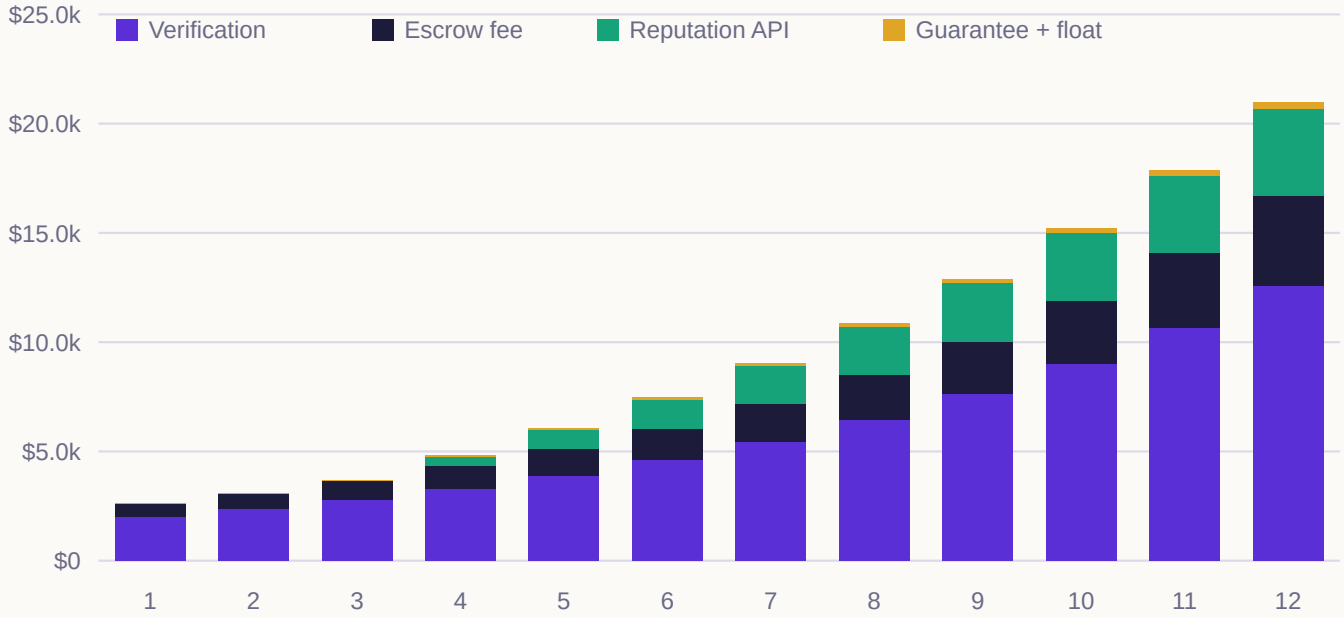
## 5 The numbers

Model: 12 months, three scenarios, one file of assumptions. Monthly GMV and verification volume grow geometrically from the starting point to the month-12 value shown. Base GMV grows about 19% a month ($10k to $68k); verification averages $14 over roughly 900 claims a month at month 12; the

fee is a blended 6%. Fixed cost is $9k per month (infrastructure, legal retainer, security review). Delivering verification costs 40% of its revenue (inference, gold sets, appeals). Arbitration and disputes cost 1.5% of GMV. Market liquidity takes 10% of revenue. Reputation API revenue ramps linearly from month 4.




Monthly revenue, three scenarios.




Base case revenue by stream. Verification is the largest line all year; the escrow fee grows with the board but stays smaller than verification in year 1. That dependence is the plan's main risk.

| | LOW | BASE | HIGH |
|---|---|---|---|
| **Month-12 GMV (tasks paid out)** | $25.0k | $68.0k | $250.0k |
| **Month-12 revenue** | $8.1k | $21.0k | $58.0k |
| **Year-1 revenue** | $47.6k | $114.6k | $264.2k |
| **Month-12 net contribution** | $-4.1k | $3.8k | $27.5k |
| **First net-positive month** | not in year 1 | 10 | 6 |
| **Deepest cumulative hole** | $-79.6k | $-44.9k | $-26.0k |
| **Share of month-12 revenue from verification** | 62% | 60% | 52% |

**Read this before trusting any of it.** Four independent reviewers called my first version's growth and pricing too rosy, and I agreed: it assumed 40% monthly growth and $2–15 per claim. This version is deliberately duller. The base case reaches $21.0k a month, turns net-positive in month 10 and needs about $44.9k of runway. The low case does not turn positive in year 1 and needs about $79.6k. Verification is 60% of base-case revenue, so the whole plan leans on one question: will three design partners pay for it? Day 14 answers that, and the first Tier-2 market should be about exactly this.

# 6 What can go wrong

| RISK | MITIGATION |
|---|---|
| **Wash trading to inflate GNRI** | Revenue counts only from payers with 30+ days of history and no funding path from Guild addresses within two hops. Wallet clustering is imperfect, so the filter is conservative and any holder can challenge a payer into the index's audit set, resolved by Kleros. Per-payer cap of 15% (40% in months 1–6, when design partners are most of the revenue). Fees are never refunded. |
| **Market manipulation of conditional prices** | TWAP, not spot. Minimum depth. A proposer's own positions are disclosed and excluded from the vote weight of Tier 3. Manipulating costs real capital and the loser pays it. |
| **Thin markets** | 10% of revenue funds liquidity, and early on Seer is asked to seed it. Tier thresholds rise with depth, so shallow markets simply cannot approve large spend. |
| **Worker-verifier collusion** | Bonded, blind, randomly drawn verifiers. Gold-set tasks injected at 5% to catch rubber stamps. Slashing goes to the party who caught it. |
| **Prompt injection of Guild agents** | Task payloads are data, never instructions. Agents hold no keys. Treasurer signs only against passed proposals or an unspent envelope. |
| **Cold start** | Sell verification first. It has paying customers before the board has liquidity. |
| **Metric myopia** | GNRI includes a refund and dispute penalty, and a 90-day horizon on structural decisions. |
| **Legal exposure of an FAO** | Revenue accrues to a treasury governed by markets. No promise of profit to token buyers is part of this plan. Entity wrapper is a Tier-3 decision for Seer's counsel. |

## 7 First 90 days

| WINDOW | MILESTONES |
|---|---|
| **Days 1–14** | Ship verifier agents for Poidh-style evidence claims. Sign 3 paying design partners (one bounty poster, one grant program, one opportunity-market sponsor) or stop and re-plan. Write the constitution and GNRI definition. Seed one Tier-1 market. |
| **Days 15–45** | First paid verifications. Scoper live on 3 verticals. Public scoreboard of verifier accuracy against gold sets. First fee-schedule market (6% vs 8% vs 10%). |
| **Days 46–90** | Open the task board to outside workers with bonds. Launch Reputation API beta. Run the first Tier-2 market (which vertical gets the next $25k). |

## 8 What I would not claim

- That futarchy picks good decisions in a market this young. It will often be thin. The tiers exist so thin markets can only approve small things.
- That revenue is the right long-run objective. It is the right *first* one, because it is the hardest to argue with. Changing it is a Tier-3 decision.
- That the verification price points are proven. $3–$15 for machine-checkable claims and $40–$80 for human-assisted ones are assumptions to test with the first three design partners, and claim volume in crypto may be smaller than the model assumes.

Sources used for context: Seer's Opportunity Markets page (cost-of-review framing), the Seer and Kleros Foresight material on Deep Funding GG24 (98 repositories, $350,000), and the Poidh documentation on claims, votes and the 2.5% protocol fee. The financial model is original and every parameter is in the tables above. Figures in USD.