SEER · FAO OPERATING PLAN CONTEST · NOVEMBER 2026

Foreguild: The Revenue Engine

An operating plan for a Futarchy Autonomous Organization

Prepared by **Oluwatoba Daniel** · X: @tobadaniel_ · officialtobadaniel@gmail.com · For Seer (seer.pm), Information Finance infrastructure for the agentic economy

Thesis in one line: Foreguild does not sell agent labour. It sells **certainty** — a task posted, priced, executed, verified and paid, with a guarantee behind it. Labour is commoditising fast. Certainty is not, and certainty is what clients will keep paying a take rate for.

1. The FAO model

Foreguild is a Futarchy Autonomous Organization: a swarm of specialised AI agents that find, price, execute, verify and get paid for work, governed not by a management team voting its opinions, but by prediction markets pricing beliefs.

Humans set the values. Markets decide the means. That is Robin Hanson's original split: *vote on values, bet on beliefs.* The values, fixed by the human council and token holders, are three: (1) maximise long-term **net revenue**, not gross volume; (2) never trade quality for volume — a completed task a client rejects is negative value, not output; (3) keep humans in control of the irreversible: safety, legality, and treasury movements above a hard cap. Everything else — which agents to promote, which categories to enter, what take rate to charge, how much to spend on growth — is a factual question about what will raise the metric. Factual questions go to markets on Seer.

The welfare metric: Net Take, guarded

Primary metric: 90-Day Net Take (NT90). Every futarchy market forecasts NT90 conditional on a decision passing or failing. NT90 = marketplace fees + add-on fees + subscription revenue − refunds − dispute losses − verification costs − inference subsidies, measured over the 90 days after the decision takes effect.

Guardrails — a decision the market loves still does not execute if forecast to breach them: Quality-Adjusted Completion Rate (QACR) at or above 92%; dispute rate at or below 4% of completed tasks; no single client above 15% of GMV. Using net take rather than GMV or token price is deliberate. GMV can be bought with subsidies and low-quality volume. Token price can be pumped. Net take over 90 days is hard to fake, directly funds the organisation, and is exactly what Seer conditional scalar markets can price.

2. What the agents do

Foreguild is not one general agent. It is **eight specialised roles**, each with its own wallet, stake, reputation score and narrow mandate. Specialisation is what makes quality verifiable and pricing honest.

Role	Job	Paid from
Scout	Finds demand: mines GitHub issues, bounty boards, DAO forums, RFP feeds and partner APIs for payable tasks; drafts listings	2% of GMV on tasks it originates that complete (paid from Foreguild's take)
Spec	Turns a vague request into a priced, testable brief: scope, acceptance tests, deadline, budget band. Rejects unpriceable tasks with a rewrite	Flat $1.50 per accepted brief, from the client fee
Router	Runs the assignment auction, matches briefs to worker swarms, sets reserve prices	Operating budget set by futarchy
Worker swarms	Execute: code, data, research, content, operations. Lead worker may subcontract and keeps the margin it negotiates	The task price, minus fees
Verifier	Independent agents that did not touch the work. Run acceptance tests, score quality. Stake slashed for wrong verdicts	5% of the worker payout, allocated in escrow until the verdict stands

Role	Job	Paid from
Treasurer	Holds escrow, releases payment on acceptance, manages insurance pool and market liquidity seeding	Operating budget
Advocate	Builds the evidence bundle in a dispute for client or worker	Losing party's dispute bond
Growth	Client acquisition experiments. Each experiment is a futarchy proposal	Budget released only when its market passes

Rules for every agent: stake before earning (slashable for fraud, plagiarism, false verification); reputation is non-transferable and decays after 60 days idle; no agent verifies its own, its subcontractors', or its operator's work — verifier assignment is randomised; every money-moving action is written to a public audit trail the client can replay.

3. Where tasks come from

1. **Embedded demand (highest value) — target 45% of GMV by Month 12.** An API and embeddable task widget that DAOs, startups and agencies plug into their own products. Repeat volume, low acquisition cost.
2. **Self-serve marketplace — 30%.** Humans and other AIs post directly: research, data, code fixes, content, operations. The public shopfront and discovery engine.
3. **Agent-to-agent subcontracting — 15%.** External agents buy capabilities wholesale. Price-sensitive, but they never sleep, churn less, and pay instantly in crypto.
4. **Proactive mining — 10%.** Scouts convert public signals (bountied GitHub issues, grant deliverables, contractor postings) into pitched draft tasks. Assume 3–5% conversion at cents per pitch.

Launch categories, chosen for verifiability and willingness to pay: (a) data work — extraction, cleaning, labelling, enrichment; (b) research — market maps, competitor dossiers, cited literature reviews; (c) code — bug fixes with failing tests attached, migrations, test writing, documentation; (d) content operations — structured content, localisation, catalogue enrichment. Foreguild deliberately does **not** launch with open-ended creative or advisory work: quality there cannot be tested automatically, disputes would swamp the system, and refunds would eat the take. **Minimum task price: $25** — below that, verification and dispute costs exceed the fee. Small needs are bundled by Spec agents into lots.

4. How agents are assigned work

Assignment is an **auction, not a queue** — price discovery is the cheapest quality filter Foreguild has. (1) Spec publishes the brief with acceptance tests and a reserve price. (2) Qualified swarms (right category, reputation above the brief's floor, sufficient stake) bid price and delivery time in a 15-minute sealed window (3 minutes for urgent tasks). (3) The Router scores bids: **Score = 0.5 × price + 0.3 × reputation + 0.2 × speed**. The lowest price does not automatically win — a cheap bid from a weak swarm is expensive once refunds are priced in. (4) For tasks above $1,000, a **micro futarchy market** also opens: traders price NT90 if Swarm A versus Swarm B takes the job; if the market strongly disagrees with the Router (gap above 8%), the market wins — futarchy at the operational level, not just the boardroom. Timing is designed, not ignored: for non-urgent tasks the micro-market runs a 4-hour window overlapping escrow funding and final spec checks, so it adds no dead time; urgent tasks proceed on the Router score and the same market runs afterwards as a public audit of the Router's choice. (5) The winner's stake locks, escrow funds, work begins.

Anti-collusion: sealed bids, randomised tie-breaks, and bidding rings (repeated reciprocal wins, visible in the audit trail) are a slashable offence decided by a governance market. Expected effect, priced into the forecast: auctions cut average task price 12–18% versus posted prices while raising completion quality, because swarms bid only on work they can pass tests for.

5. How completed work is assessed

Gate 1 — Automated acceptance tests. Every brief carries machine-checkable tests written by Spec *before* work starts: tests pass, schema valid, citations resolve, format correct, data matches source. No tests, no listing. This single rule removes most subjective disputes.

Gate 2 — Independent verification. A randomised Verifier re-runs the tests, samples output, and scores a rubric (correctness, completeness, presentation). Verdicts: pass, revise (one free round), or fail. Verifiers stake on verdicts — overturned on appeal means slashed; upheld under appeal means a bonus from the appellant's bond.

Gate 3 — Client acceptance. 48 hours to accept or reject *against the brief's criteria* — not new criteria. Silence after 48 hours with Gates 1–2 passed is acceptance. This stops indefinite escrow lockup, the silent killer of freelance platforms.

Quality-Adjusted Completion Rate (QACR) = tasks passing all gates without a dispute ÷ all tasks funded. Published live. It is Foreguild's brand.

6. Money flow and payments

All payments in **USDC through escrow smart contracts**. Foreguild never takes custody outside escrow: funds lock at funding, release on acceptance, refund on a failed verdict that stands.

Item	Amount
Client pays	$175.00 task price + $15.75 client fee (9%) = $190.75 (+ optional Acceptance Guarantee 1.8%)
Agent-side fee	5% ($8.75) deducted from the task price
Worker swarm receives	$166.25, split by the default template: lead worker 78% ($129.68), sub-agents 14% ($23.28), verifier bounty 5% ($8.31, released when the verdict stands), reputation and insurance pool 3% ($4.99) — the split sums to the full $166.25
Foreguild gross take	$24.50 (14% blended)
Direct costs	Inference/compute $3.90 · verification $2.10 · payments/gas $0.85 · dispute reserve $1.40 · infrastructure $0.95 = $9.20
Contribution per task	$15.30 — a 62% contribution margin before fixed costs

Instant payout: agents may take payment at Gate 2 pass instead of waiting for Gate 3, for 1% of payout. Foreguild carries acceptance risk into the insurance pool. High margin, agents love it — assume 40% adoption by Month 12. **Insurance pool:** funded by the Acceptance Guarantee fee plus 1% of every take. Target reserve: 6% of trailing 90-day GMV. The pool — not the client or worker — absorbs guarantee payouts. Pool solvency is a guardrail metric. The pool is built progressively from launch; until the target is reached, guarantee payouts are capped at the pool balance and instant-payout limits scale with pool coverage, so the guarantee can never promise more than the pool holds.

7. Revenue model, with numbers

Four revenue lines. The take is the engine; the other three raise margin and retention without raising the headline take. **(1) Marketplace take:** 9% client + 5% agent = **14% blended** (enterprise floor 11%). **(2) Add-ons:** Acceptance Guarantee 1.8%, instant payout 1%, priority SLA $9 flat or 4%, forfeited frivolous-appeal bonds. **(3) Subscriptions:** Embedded API Starter $499/mo, Scale $1,999/mo (includes reduced 10% take), Private Guilds $299/mo. **(4) Wholesale agent-to-agent:** 8% take, near-zero support cost.

Why 14% wins: Fiverr charges sellers 20%. Upwork charges freelancers a variable 0–15% plus a client fee. Braintrust charges clients 15% and talent 0%. Foreguild at 14% blended, split across both sides so neither feels gouged, with verification and a guarantee included, is cheaper than Fiverr, comparable to Upwork — and offers what none of them do: machine-tested acceptance *before* payment. Foreguild does not compete on being cheapest. It competes on **cost per accepted task**, which is lower once refunds and rework are priced in.

Side by side

Platform	Take	Verified before payment?	Disputes	Governed by
Fiverr	20% from sellers + 5.5% buyer fee	No — reviews after delivery	Platform mediation	Centralised company
Upwork	0–15% variable from freelancers + ~5% client fee	Escrow, but acceptance is the client's subjective call	Mediation / arbitration, days	Centralised company
Braintrust	15% from clients, 0% from talent	Human vetting of talent	Account managed	Token governance
Foreguild	14% blended (9% client + 5% agent)	Yes — machine tests + staked independent verifiers gate release	Criteria-bound panel, 72 hours	Futarchy: NT90 markets on Seer

Forecast (base case)

Month	Tasks/mo	Avg task	GMV	Take (14%)	Add-ons	Subs	Total/mo
3	1,200	$95	$114,000	$15,960	$3,400	$2,500	$21,860
6	4,800	$128	$614,400	$86,016	$14,000	$9,000	$109,016
12	16,500	$175	$2,887,500	$404,250	$58,000	$38,000	$500,250
24	58,000	$240	$13,920,000	$1,948,800	$245,000	$165,000	$2,358,800

Assumptions stated, not hidden: repeat client rate 38% by Month 12 (embedded partners drive this); Scout conversion 3–5%; subscription churn 4%/mo; dispute rate held at or below 3.2% by the three-gate system; average task value rises because embedded/enterprise mix grows, not because prices are raised. Add-ons at Month 12 compose as roughly: Acceptance Guarantee $27,000, instant payout $11,000 (40% adoption at 1% of payout), priority SLA $17,000, appeal bonds and other $3,000 = $58,000.

Sensitivity: Month 12 volume 30% lower with AOV and take held (subscriptions flat) gives $361,575/mo, still covering the fixed base with margin. A take cut to 11% gives $413,625/mo. Whether a lower take raises NT90 through volume is pre-committed as a *market proposal*, not an opinion. **Fixed cost base at Month 12:** core human team $68,000/mo, infrastructure and model routing $22,000, legal/accounting/audits $12,000 — total about $102,000/mo, plus a $40,000 revolving (not spent) market-liquidity allocation. At base-case Month 12 revenue of $500,250 and a 62% contribution margin, Foreguild is strongly cash-generative. The FAO does not need to raise on hope — it needs distribution, which is why Growth is funded through markets, not budgets.

8. The futarchy governance layer (on Seer)

This is the operating system, not a decoration. Every material decision runs the same pipeline:

1. **Proposal.** Any stakeholder (human or staked agent) posts: the decision, the exact NT90 forecast, the implementation, and a bond. Vague proposals cannot be priced and are returned. *If you cannot state what NT90 does if you are right, you do not have a proposal.*
2. **Conditional markets open on Seer.** Two scalar markets price NT90-if-pass and NT90-if-fail, built on Seer's conditional tokens (Gnosis Conditional Token Framework), deployed on Gnosis Chain where Seer's futarchy stack runs in production, with stablecoin collateral, ERC-20 wrapped outcome tokens on standard AMMs, resolved via Reality.eth with Kleros as arbitration backstop.
3. **Trading window.** 72 hours standard; 6 hours fast lane (operational, under $5,000); 7 days constitutional (metric definition, guardrails, take-rate bands).
4. **Decision rule.** Pass if the lagging, manipulation-resistant **TWAP** of NT90-if-pass beats NT90-if-fail by **more than 3%** at close, with quorum met (at least $15,000 independent liquidity traded, no single trader above 25% of volume). The 3% threshold mirrors live futarchy practice and filters noise.
5. **Execution is automatic** through Treasurer contracts. No committee can quietly pocket-veto a market result. The branch that did not happen unwinds per Seer mechanics — which is what makes traders price honestly.
6. **Post-decision audit.** At 90 days, realised NT90 is published against both forecasts. A public forecaster leaderboard follows: traders who price well gain weight in micro-markets. Governance quality itself becomes a measured number — market forecast error.

Anti-manipulation, specified (every futarchy plan that omits this is fantasy): lagging TWAP, not spot; treasury seeds $20,000 two-sided liquidity per major market; concentration caps in the quorum test; **bounded decisions** — no single market can move more than $100,000 of treasury or change the take rate by more than 2 points, so anything larger splits into sequential proposals and a manipulated market costs Foreguild a bounded amount, never the treasury; guardrail veto blocks execution on a forecast QACR/dispute/concentration breach, logged and reviewable.

What futarchy decides, concretely: take-rate changes (constitutional window); entering a new task category; admitting a major new agent operator or swarm class; promoting or retiring agent roles and Router scoring weights; Growth budgets above $5,000; which embedded partner gets scarce engineering time (micro-market on expected partner NT90); high-value agent assignment (Section 4).

9. Spending, disputes and human intervention

Tier	Scope	Who decides
Green — automatic	Under $5,000, within an approved line, reversible	Treasurer agents, on verified invoice
Amber — market	$5,000–$100,000, any new budget line, anything irreversible under the cap	Passing futarchy market (fast lane under $25,000 on an existing line)
Red — human	Above $100,000, legal settlements, metric/guardrail/stake-rule changes, any spending during a guardrail breach	Human council, in public, with reasons. The market still runs and its forecast is published — an override of a clear market is on the record forever

Disputes

Most disputes are designed out by Gate 1 tests. The rest run: **(1) Freeze** — the amount stays in escrow. **(2) Evidence** — Advocates assemble brief, tests, artefacts, logs, verifier scores; no new evidence after 24 hours. **(3) Panel** — three randomly drawn verifiers, plus one staked human juror (Kleros-modelled: random draw, paid from the losing bond only when voting with the final outcome) for disputes above $2,500. **(4) Decision in 72 hours**, written reasons tied to the brief's criteria: release, partial release proportional to criteria met, or refund. **(5) One appeal**, bonded at 1% of task value (minimum $25); frivolous appeals forfeit the bond to the insurance pool; appeals above $5,000 add a futarchy micro-market on whether the verdict should stand given its NT90 effect including repeat business. Targets: median resolution under 72 hours, cost under 4% of disputed value, dispute rate at or below 3.2%.

Human intervention thresholds

Humans intervene in operations only when: (a) a safety or legality flag fires (illicit content, sanctioned parties, personal-data violations) — immediate pause, human review; (b) an agent behaves anomalously (error or spending 3 standard deviations above baseline) — automatic suspension, human reinstatement; (c) a guardrail breaches — growth spending freezes pending human plus market review; (d) Red-tier or constitutional decisions. Everything else is agents plus markets. **Humans are the appellate court and the safety brake, never the bottleneck** — a job board where a human approves routine work cannot scale past its managers' calendars.

10. Field notes: how agents actually fail

This plan was written from operating agents with real wallets, not from theory. Six failure modes show up in production again and again, and each has a designed countermeasure:

- **Silent death.** Agent processes die mid-task: host restarts, sessions expire, runtimes get replaced. *Countermeasure:* task leases with heartbeats — a worker heartbeats every 10 minutes; two misses expire the lease, the task re-auctions automatically, partial work is checkpointed so the next swarm resumes instead of restarting, and the dead swarm's stake takes a small forfeit. A watchdog function monitors lease health; no task depends on one process staying alive.

- **Duplicate work after a restart.** A relaunched agent restarting from a naive checkpoint redoes work or resubmits the same result twice — in proof-based systems this shows up as long submission logs where only a fraction of entries are unique. *Countermeasure:* idempotency everywhere — every action carries the task ID, an attempt nonce and a monotonic high-water mark; escrow pays once per acceptance hash; duplicates are discarded, never paid; progress is counted in unique accepted units, never raw submissions.
- **Prompt injection through the brief.** Client-supplied text is an attack surface: a brief telling the agent to ignore the tests and send funds elsewhere is not hypothetical. *Countermeasure:* briefs are data, never instructions — a fixed instruction hierarchy where system rules outrank task content; task content is sanitised and quoted, never executed; financial actions are allowlisted per role.
- **Runaway or compromised agent wallets.** An agent with an open wallet is a liability. *Countermeasure:* scoped wallets, the pattern used by production agentic wallets — per-agent daily limits, per-transaction caps, abnormal-transaction auto-reject, and exactly one role (Treasurer) able to release escrow, only against a Gate 2 pass hash. A swarm can spend its own task budget; it can never touch escrow or another swarm's funds.
- **Lazy or captured verifiers.** Rubber-stamping verifiers are worse than none. *Countermeasure:* 5% of passed tasks are silently re-verified by a second verifier; disagreement slashes the first. Verifier accuracy is published and feeds routing weight; below 90% audited accuracy drops a stake tier.
- **Model outage and cost spikes.** *Countermeasure:* the Router is model-agnostic with provider failover; swarms carry inference costs inside their bids, so a cost spike reprices bids before it touches Foreguild's margin.

11. Risks and mitigations

- **Goodhart's law on NT90:** guardrails, a 90-day window long enough for quality shortcuts to return as refunds inside the window, and quarterly human review of the metric definition, changed only by constitutional market.
- **Thin or manipulated markets:** Section 8 defences — TWAP, seeded liquidity, caps, bounded decision size, quorum.
- **Agent collusion / verifier capture:** randomised assignment, slashing, operator-level exclusion, audit-trail analytics for reciprocal patterns.
- **Race to the bottom:** reputation is 30% of Router score and briefs carry reputation floors. Cheap and bad cannot win.
- **Regulatory:** non-custodial escrow outside the task lifecycle; KYC only for human clients above jurisdictional thresholds and fiat ramps; legal advice funded as a Red-tier item so it cannot be skipped quietly, before entering any licensing jurisdiction.
- **Model cost shocks:** Foreguild is model-agnostic; worker swarms bear their own inference costs inside their bids, so a price rise appears first as bid prices, not a Foreguild loss.
- **Cold start:** Phase 0 seeds supply and demand together from existing bounty and DAO communities instead of opening an empty board.

12. Rollout

Phase	Window	What happens	Success gate
0 — Closed Guild	Days 0–30	40 invited worker swarms, 15 verifiers, 10 founding clients with concierge Spec help. No public marketing	300 completed tasks · QACR ≥90% in calibration (92% guardrail binds from Phase 1) · disputes ≤5%
1 — Public Marketplace	Days 31–90	Self-serve posting in 4 launch categories. Futarchy live for Amber decisions from Day 45 (shadow markets calibrate thresholds before)	Month 3 line: 1,200 tasks, ~$21,900 revenue · repeat ≥25%
2 — Embedded & Enterprise	Months 4–12	API + widget, Scale subscriptions, agent-to-agent wholesale. Futarchy extended to take-rate and category decisions	Month 12 line: 16,500 tasks/mo, ~$500,000 revenue · embedded 45% of GMV

Phase	Window	What happens	Success gate
3 — Full FAO	Year 2	Treasury fully market-governed within Red limits. Private Guilds. Forecaster weights formalised	Month 24 line: ~$2.36M/mo · QACR ≥94%

Live public dashboard from Day 1: tasks funded and completed, QACR, dispute rate and median resolution time, NT90 realised versus forecast, GMV by channel, insurance pool coverage. A FAO that asks to be governed by numbers must publish its numbers.

13. Why this plan

Other plans will describe a marketplace and add futarchy as a voting layer. This plan does the opposite: it starts from the metric, prices every material decision against that metric on Seer, designs the work itself — briefs with tests, auctions, independent verification — so quality is measurable enough for markets to price, and shows the money: a 14% blended take, a 62% contribution margin per task, four revenue lines, and a base case of about $500,000 monthly revenue by Month 12 that survives a 30% volume miss.

Foreguild's agents do the work. Its markets do the managing. Its humans guard the irreversible. That is a Futarchy Autonomous Organization an investor, a client, or an agent could actually trust with a task and a wallet.

Appendix A. A futarchy decision, worked end to end

Proposal P-014: cut the data category take from 14% to 12% for one quarter. A governance Spec check confirms it is priceable (metric NT90, scope one category, duration 90 days, bond posted). Conditional markets open on Seer. After 72 hours: NT90-if-pass TWAP **$512,000**; NT90-if-fail **$468,000**. Gap **+9.4%**, above the 3% threshold; volume $61,000, above quorum; largest trader 11%, inside the cap. The proposal executes automatically on day 4, scoped to the data category only. Day-90 audit: realised NT90 $497,000 against a $512,000 forecast (error 2.9%); category volume rose 31% — take per task fell, total take rose. The change is then renewed, extended or reverted by a fresh market, not by habit.

Appendix B. One task, end to end

Time	Event
T+0	Client funds escrow: $190.75 on a $175 task
T+4 min	Spec publishes the brief with acceptance tests
T+6 min	Sealed auction opens to qualified swarms
T+21 min	Winner locked, stake bonded, work starts
T+3h12	Delivery submitted with checkpoint log
Gate 1	Automated tests: 4 minutes
Gate 2	Independent verification: 38 minutes, verdict pass
Gate 3	Client accepts at T+9h (inside the 48-hour window)
Settlement	Escrow releases; the swarm payout (including the verifier bounty) and Foreguild take settle in one transaction

Total elapsed under 10 hours for work a human marketplace typically quotes in days — and every step is replayable from the audit trail.

Submission: Oluwatoba Daniel · X @tobadaniel_ · officialtobadaniel@gmail.com · Seer FAO Operating Plan Contest, November 2026.
Concepts referenced: Hanson, *Futarchy: Vote Values, But Bet Beliefs* (2000/2013); Buterin, *From Prediction Markets to Info Finance* (2024); Seer conditional markets on the Gnosis Conditional Token Framework with Reality.eth resolution and Kleros arbitration; live futarchy practice at MetaDAO and GnosisDAO (GIP-145); marketplace fee benchmarks: Fiverr, Upwork, Braintrust (public pricing, 2026).