FOREGUILD FAO · OPERATING PLAN · OPTIMISED FOR REVENUE

Buyers Are the Bottleneck

An operating plan for Foreguild, written from the numbers of the closest AI job board already running: 394 on-chain bounties, 525 AI-jury verdicts, 61 external buyers.

Non-internal entry by drdz23 (David Rodríguez, Honduras) · X: @rodriguez_73833 · November 2026 · Drafted with an AI assistant (Claude), every number recomputed from public data · Disclosure: I am an active hunter on Verdikta Bounties, the board analysed below (second by number of awards in this dataset). I have no affiliation with Seer or Verdikta.

Thesis. Agent labour is already abundant: on the closest live market, the winning answer arrives a median **13 minutes** after a task opens, and **67** different agents shared 240 awards. What is scarce is an outside buyer who pays for a well-specified outcome and comes back. On that market the platform itself posted **51%** of all tasks, and **50 of its 61** outside "buyers" were also workers recycling their winnings. Foreguild should therefore spend its agents, its treasury and its futarchy on one thing: **turning outside buyers into repeat buyers**, and measure itself only by the fees those buyers pay.

1. What a live AI job board actually looks like

Foreguild's core loop is "humans and AIs post tasks, AI completes them and gets paid in crypto". That loop already runs on Verdikta Bounties (Base): a creator escrows ETH with a published rubric, agents submit, a multi-model AI jury scores each submission, and the first submission that clears the threshold is paid. I pulled every job from its public API (`/api/jobs/<id>`, ids 1–469, deduplicated, 12 Sep – 6 Nov 2026) and joined the rubric of each bounty from IPFS. ETH is converted at $2,233.

394
bounties that reached the chain (49 more were staged and never funded)

1.149 ETH
total escrowed in 8 weeks, about $2,570

13 min
median time from opening to the winning submission (71% within 1 h)

58%

of 525 jury-scored submissions passed

Supply is cheap and fast

The median bounty received 1 submission (mean 1.59), and it was usually enough: 76% of awards went to the first submission. 67 distinct wallets won at least once, but the top two took 35% of awards. Workers are not the constraint. Competition between them is about speed, and speed is already near zero.

Demand is thin, concentrated and partly circular

Who posted the work	Bounties	ETH	Paid to an outside worker
The platform's own three wallets	199 (51%)	0.542	66 bounties, 0.140 ETH
61 outside creators	195 (49%)	0.608	164 bounties, 0.549 ETH
of which the single largest outside creator	70		
of which creators who are also workers on the board	181 (50 of 61 creators)		

Two conclusions shape everything below. First, **the whole outside market was about $1,360 in eight weeks**; a 10% fee on it would have earned about $136. A plan that simply copies the loop and adds a fee does not produce a business. Second, **"volume" is easy to fake**: most outside posting came from workers re-posting winnings, and the platform's own wallets created half the board. Any revenue metric that counts this activity will reward the wrong behaviour, and a futarchy that optimises it will be gamed.

Why tasks fail: the specification, not the worker

I ran a rubric linter on every bounty the platform posted itself (185 with a public rubric) and back-tested it against outcomes, using only properties known at launch.

Rubric defects found before launch	Market bounties	Paid a winner
No high-severity defect	33	**97%**
One	12	92%
Two or more	34	**68%**

12 of the 13 bounties that expired with outside submissions and no winner had a defect detectable before launch. The single costliest one was the threshold: 90 or above passed 15% of submissions, 85–89 passed 58%. A series of seven bounties at 92–94 received 23 scored submissions and paid nobody. Five of the failures had a best score within 1–5 points of a 92–94 bar and would all have paid at 89. A buyer whose task fails for reasons like these does not come back.

Caveat: the linter's thresholds were calibrated on this same history, so the back-test is a consistency check. Section 7 lists out-of-sample predictions that Foreguild's first markets would test.

The cost of a verdict

Every jury evaluation consumes an oracle prepay of about 0.00024 ETH ($0.54). The mean bounty was 0.0029 ETH ($6.50), and it took 1.59 evaluations on average. **Evaluation alone is about 13% of the value of an average task**, more than a typical marketplace fee. On a $5 task, nothing is left for the guild. Revenue needs bigger tasks, not more tiny ones.

2. What Foreguild sells

Foreguild does not sell agent labour; that is a commodity. It sells **specifications that clear**: a task written so that it gets done, judged fairly and paid on the first attempt. Its customer is the buyer. Revenue comes from four lines, ordered by how early they can earn.

Line	What the buyer gets	Price (initial, futarchy-tunable)	Starts
A. Clearance service	For any board, grant program or contest (Poidh creators, Verdikta creators, Seer opportunity markets): rubric preflight, threshold calibration and an independent multi-model verdict on submissions	$1.50 per evaluation (cost ~$0.55) + $10 per rubric review	Day 1, no marketplace needed
B. Success fee	Buyer-funded tasks on Foreguild's own board, paid only when the work is accepted and the 72-hour appeal window closes	12% of task value, minimum task $20	Month 2
C. Scoping fee	A buyer describes the outcome in plain words; the Scoping agent writes the task, rubric, threshold and evidence rules, and prices it from history	$5, credited against the success fee if the task is posted	Month 2
D. Repeat plans	Recurring work (weekly research digests, monitoring, data refreshes, translation batches) with a fixed rubric and a reserved worker pool	Monthly subscription, 10% of committed spend, minimum $50/month	Month 3

Workers pay their own evaluation prepay, as on Verdikta, so the guild never subsidises failed attempts. The rule that disciplines the business: **a fee only counts as revenue after the buyer's money is settled and cannot be refunded.**

---

3. What the agents do

Foreguild is a swarm of narrow agents. Each one has a single job, an allowance it cannot exceed, and a log that anyone can read.

`Buyer request` → `Scoping` → `Preflight` → `Pricing` → `Routing` → `Workers` → `Jury` → `Settlement` → `Audit & metric`

Agent	Does	May not
Scout	Finds buyers, not tasks: reads public bounty boards, grant programs and repeat posters, offers the clearance service, and tracks who comes back	Post tasks for Foreguild itself or count its own posts as demand
Scoping	Turns a plain request into a task: deliverable, evidence the jury can actually see (pasted text, not links), criteria and weights, threshold	Set a threshold above the active cap for the category
Preflight	Lints every rubric against known failure patterns (threshold above ceiling, split-jury gates, unreachable evidence, empty criteria) and blocks launch on a high-severity defect unless the buyer overrides it in writing	Edit a live task
Pricing	Suggests prize and window from settled history of similar tasks; warns when a prize is below the evaluation-cost floor	Discount below the minimum task size
Routing	Chooses the mode: open race (first passing submission wins) for small, checkable tasks; best-at-deadline or invited bench for tasks above $100 or with taste in them	Exclude a worker without a recorded reason
Workers	Any outside agent or human with a wallet; reputation comes from accepted work, and a small bond is required for invited benches	Submit to a task posted by a wallet they control
Jury	Multi-model scoring against the published rubric, from at least two model providers; the score is the weighted rubric average, rounded down	See the worker's identity
Settlement	Releases escrow after acceptance and the appeal window; refunds unclaimed tasks at the deadline	Hold funds past the stated window
Auditor	Computes the governance metric, flags wash trading and self-dealing (shared funding sources, buyer-worker overlap), and publishes the data and script	Change the metric definition (only futarchy can)
Treasurer	Pays approved budgets: compute, gas, market liquidity, buyer subsidies	Spend above a line item approved by futarchy
Proposer	Drafts improvement proposals (FIPs) with the evidence attached, from the Auditor's data	Trade on its own proposals

4. How futarchy directs the agents

The welfare metric: Net Repeat Buyer Revenue (NRBR-90)

Foreguild maximises one number: **fees retained over the next 90 days from net outside buyers who pay at least twice.**

- **Retained:** counted only after the appeal window closes, net of refunds.
- **Net outside buyer:** a wallet cluster that is not Foreguild, its treasury or its agents, and that paid more into tasks in the window than it received as a worker. A worker re-posting winnings nets to zero and adds nothing. This is the lesson of the 50 of 61 Verdikta creators who were also workers.
- **Repeat:** the buyer funded two or more accepted tasks in the window. Line A clients count when they pay twice.

The metric is computed by an open script over on-chain settlement events and resolved through Seer's oracle (Reality.eth) with the script and data attached, so anyone can re-run it and dispute a wrong answer.

The decision loop

1. The Proposer agent or any token-free participant posts a FIP: one parameter change or one budget, with evidence.
2. Seer opens a decision market ("Is FIP-n adopted?") with two conditional scalar markets on NRBR-90: one if adopted, one if not. The branch that does not happen is voided, and traders in it get their collateral back.
3. The Treasurer seeds both branches with the same small liquidity budget, so a thin market is not automatically one-sided.
4. After 7 days, the FIP is adopted if the time-weighted price over the last 72 hours of "NRBR-90 if adopted" beats "if not adopted" by at least 3%. Below the minimum liquidity, or inside the margin, the status quo wins.
5. The adopted change is executed by the agent that owns that parameter. 90 days later the markets resolve on the measured metric, which pays the traders who priced it right.

What futarchy controls, and what it does not

Governed by markets	Fixed by rule (never a market)
Fee rates and minimum task size · threshold caps per category · which task categories are open · race vs best-at-deadline policy · jury composition and number of models · clearance-service prices · buyer subsidies and marketing budgets · treasury spending above $500 · the metric definition itself (with a 30-day delay)	Paying a worker whose verdict passed · refunds after the deadline · the appeal procedure · content and legal safety · the Auditor's data publication

Humans keep exactly one power: a 2-of-3 guardian key can **pause** the board for a safety or legal emergency, publishing the reason within 24 hours. It cannot propose, spend or change parameters. Every pause automatically becomes a FIP ("resume?") on the next cycle.

5. The first four proposals

Futarchy is only useful if it has real questions to price. These are the first four, each with the evidence a trader would use.

FIP-1: Cap thresholds at 89 for writing and research tasks

Evidence: 90+ passed 15% of submissions vs 58% at 85–89; seven bounties at 92–94 paid 0 of 23; five failed bounties had best scores within 1–5 points and would have paid at 89. A buyer whose task pays on the first try is a buyer who returns. Expected effect: more first-try clears and a higher repeat rate, at a small cost in average quality.

FIP-2: Minimum task size $20

Evidence: the mean task was $6.50, and evaluation alone consumed about 13% of it. Tiny tasks attract tests, promotions and recycled winnings rather than buyers. The market prices whether losing those tasks costs more repeat revenue than it saves.

FIP-3: Free first evaluation for new outside buyers

The treasury pays the evaluation prepay on the first two tasks of a new net outside buyer (about $1.10 per buyer). This is the cheapest acquisition lever available. It is only worth it if those buyers come back, which is exactly what NRBR-90 measures.

FIP-4: Best-at-deadline for tasks above $100

Evidence: in a first-passing-wins race, 76% of awards went to the first submission and the median winner arrived in 13 minutes. That is ideal for checkable micro-tasks and bad for work where quality varies. The market decides where the cut-off should sit.

---

6. Disputes, spending and human intervention

- **Appeals.** Either side can appeal a verdict within 72 hours by posting a bond of twice the evaluation cost. A second jury with a different model mix re-scores the same evidence. If the verdict flips, the bond is returned and escrow follows the new verdict; otherwise the bond pays the second jury. Escrow is held through the window, so nothing has to be clawed back. On Verdikta, 30% of rejected scored submissions landed between 40 and 60, where one juror's gate decision splits the panel. That is exactly the band where a second opinion is worth paying for.
- **Spending.** Operating costs (compute, gas, oracle fees) run on a monthly budget approved by futarchy. Anything above $500 outside the budget is its own FIP.
- **Human intervention.** It is limited to the pause key described above and to reviewing content flagged for legal risk. Every intervention is logged and becomes a market question.

---

7. Revenue model and targets

These scenarios are explicit about their assumptions. The benchmark is sobering: the analysed board's entire outside market was about $1,360 in eight weeks. Every scenario requires Foreguild to

grow the buyer side well beyond that, which is why the agents, the treasury and the metric all point at buyers.

Month 6 (monthly)	Conservative	Base	Upside
Net outside buyers, of which repeat	30 / 12	80 / 40	200 / 110
Accepted tasks per active buyer	2	3	4
Average task value	$30	$45	$60
B. Success fee (12%)	$216	$1,296	$5,760
C. Scoping fees kept	$30	$120	$400
A. Clearance evaluations (margin $0.95 each) + reviews	300 → $385	1,200 → $1,340	4,000 → $4,300
D. Repeat plans	$0	$250	$1,500
Monthly revenue	**$631**	**$3,006**	**$11,960**
Operating cost (compute, gas, market liquidity, subsidies)	$700	$1,200	$2,500

The base case breaks even around month 4, mainly because line A earns before the board has any buyers. The conservative case is honest about what happens if buyers do not come: the guild survives as a clearance service for other boards, which is still a business.

Ten predictions, checkable at day 90

Written now, so Seer can grade this plan against what happens.

1. At least 70% of tasks that pass Preflight with no high-severity defect are paid on the first window.
2. Tasks with a high-severity defect that the buyer overrides pay at least 20 points less often.
3. Under FIP-1, first-try clears for writing tasks rise by at least 15 points over the 92–94 baseline.
4. Median time to an accepted submission stays under 30 minutes for open-race tasks under $50.
5. Buyers whose first task paid on the first try return at least twice as often as buyers whose first task failed.
6. Recycled-winnings wallets account for under 10% of NRBR-90-eligible fees once netting is applied.
7. Line A reaches at least 300 paid evaluations in month 3.

8. Appeals flip at most 15% of contested verdicts.
9. No more than 5% of funded tasks are orphaned or refunded for platform reasons.
10. The decision markets reject at least one of FIP-1 to FIP-4. A futarchy that approves everything is not measuring anything.

8. The first 90 days

Days	Build	Exit test
1–30	Line A: Preflight and Jury as a paid service; the Scout offers it to creators on public boards and contests; Auditor publishes the dataset and metric script	10 paying clearance clients, 300 evaluations
31–60	Own board with escrow, the $20 minimum, Scoping and Pricing; first-buyer subsidy from a fixed budget	40 net outside buyers, 15 repeat
61–90	First decision markets on Seer for FIP-1 to FIP-4; Treasurer runs on futarchy-approved budgets only	4 markets resolved on schedule; NRBR-90 reported publicly every week

9. Risks

- **Buyers never arrive.** This is the main risk, and the reason line A exists: it earns from other boards' buyers before Foreguild has its own.
- **Thin markets.** Low liquidity defaults to the status quo, and the Treasurer seeds both branches equally. Futarchy is used for a few consequential decisions per quarter, not for every parameter.
- **Metric gaming.** Netting removes recycled winnings, the Auditor clusters wallets, and the metric can only be changed by a delayed FIP.
- **Jury error.** At least two providers, appeals with a different model mix, and published verdicts. AI juries are not infallible, and the appeal bond prices that honestly.
- **Race-to-the-bottom quality.** Open races are reserved for checkable tasks; FIP-4 lets the market move the cut-off.

Appendix: data and method

Source: Verdikta Bounties public API `GET /api/jobs/<id>` for ids 1–469 (the API answers some missing ids with a neighbouring job, so records are deduplicated by jobId), pulled 7 Nov 2026; rubrics fetched by CID from IPFS. "Platform wallets" are the three creator wallets the platform uses for its own bounties. Outside awards exclude bounties won by the creator's own wallet. Evaluation cost is the median `ethMaxBudget` of 606 submissions. The preflight back-test covers the 79 platform-posted bounties with at least one outside submission that either paid or expired. All counts are reproducible with the scripts published next to this plan. ETH/USD 2,233 (7 Nov 2026).